

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 20:21:58 ; Search time 23.68 Seconds
(without alignments)
2374.297 Million cell updates/sec

Title: US-09-787-737-2

Perfect score: 1751

Sequence: 1 MSSSNKNWPSMFKSKPCNNN.....TDEFGVTYQPLQNGAIYYLI 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriapi:*

17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1347	76.9	253	10 Q9FNM4	Q9fnn4 arabidopsis
2	467	26.7	321	10 Q23662	Q23662 arabidopsis
3	265	15.1	219	10 Q9FN71	Q9fn71 arabidopsis
4	242.5	13.8	199	10 Q9SS66	Q9ss66 arabidopsis
5	238.5	13.6	122	10 Q9FNM1	Q9fnn1 arabidopsis
6	217.5	12.4	295	10 Q9ZVF5	Q9zvf5 arabidopsis
7	199.5	11.4	291	10 Q9SB92	Q9sb92 arabidopsis
8	188.5	10.8	350	10 Q9LVH2	Q9lvh2 arabidopsis
9	188	10.7	293	10 Q9SL51	Q9sl51 arabidopsis
10	177.5	10.1	165	10 Q9C775	Q9c775 arabidopsis
11	175	10.0	197	10 Q9LM83	Q9lm83 arabidopsis
12	173.5	9.9	262	10 Q9FSR5	Q9fsr5 oryza sativ
13	171.5	9.8	122	10 Q9FFK0	Q9ffk0 arabidopsis
14	171.5	9.8	211	10 Q9LM84	Q9lm84 arabidopsis
15	170	9.7	261	10 Q9ZR24	Q9zrz4 populus tre
16	168.5	9.6	244	10 Q9S1B4	Q9s1b4 arabidopsis

17	168	9.6	250	10 Q94DV1	Q94dv1 oryza sativ
18	167.5	9.6	217	10 Q9Z825	Q9zrz5 populus tre
19	167	9.5	268	10 Q81788	Q81788 arabidopsis
20	166.5	9.5	249	10 Q9C632	Q9c632 arabidopsis
21	161	9.2	275	10 Q9LIX7	Q9lix7 oryza sativ
22	115	6.6	800	5 Q43988	Q43988 dictyosteli
23	111.5	6.4	745	10 P93041	P93041 arabidopsis
24	110.5	6.3	215	10 Q9FIE5	Q9fie5 arabidopsis
25	109	6.2	801	10 Q9SWZ6	Q9swz6 arabidopsis
26	106	6.1	731	2 Q9KKE8	Q9kke8 streptococc
27	106	6.1	3097	5 Q61143	Q61143 plasmodium
28	105.5	6.0	800	5 Q9G902	Q9g902 dictyosteli
29	104.5	6.0	276	5 Q9U739	Q9u739 podocoryne
30	104.5	6.0	956	3 Q94717	Q94717 schizosacch
31	101.5	5.8	659	10 Q41074	Q41074 phaeodactyl
32	98	5.6	468	5 Q44349	Q44349 funiculina
33	97	5.5	251	10 Q9LXG6	Q9lxg6 arabidopsis
34	97	5.5	259	10 Q9ARE2	Q9arc2 flaveria bi
35	96.5	5.5	1281	11 Q60713	Q60713 mus musculu
36	96.5	5.5	1281	11 Q91289	Q91289 mus musculu
37	96.5	5.5	1281	11 Q91288	Q91288 mus musculu
38	96.5	5.5	1300	11 Q61785	Q61785 mus musculu
39	96.5	5.5	1458	11 Q91W10	Q91w10 mus musculu
40	96.5	5.5	1487	11 Q9JLA1	Q9jla1 mus musculu
41	96	5.5	322	5 Q45024	Q45024 branchiosto
42	96	5.5	873	11 P70121	P70121 mus musculu
43	95	5.4	420	16 Q97K22	Q97k22 clostridium
44	95	5.4	938	3 Q14100	Q14100 schizosacch
45	95	5.4	1072	16 Q92X83	Q92x83 rhizobium m

ALIGNMENTS

RESULT 1

Q9FNM4 ID Q9FNM4 PRELIMINARY: PRT: 253 AA.
AC Q9FNM4;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE SIMILARITY TO HOMEODOMAIN TRANSCRIPTION FACTOR.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006698; BAB08243.1; .
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
SQ SEQUENCE 253 AA; 28442 MW; AB51C7A8B8A31A02 CRC64;

Query Match 76.9%; Score 1347; DB 10; Length 253;
Best Local Similarity 99.6%; Pred. No. 5.3e-107;
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSSNKNWPSMFKSKPCNNHHQHHEIDTPSYMHYSNCNLSSFSFSDRIDPDKPRWPKP 60

Db 1 MSSSNKNWPSMFKSKPCNNHHQHHEIDTPSYMHYSNCNLSSFSFSDRIDPDKPRWPKP 60


```
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DE EMBL; AC009327; AF03478.1; -.
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 199 AA; 22314 MW; 1C8EB9FA32C12661 CRC64;

Query Match 13.8%; Score 242.5; DB 10; Length 199;
Best Local Similarity 53.9%; Pred. No. 6e-13;
Matches 48; Conservative 15; Mismatches 25; Indels 1; Gaps 1;

QY 29 TP-SYMHYSCNLSFSSDRIPDPKPNPKPQIIRLESIFNSGTINPPREIQRIRI 87
Db 6 TPSPTRHSRPPSSAGSSTAEPVRWSKPKPQIIRLESIFNSGTINPPREIQRIRI 65

QY 88 RLQRYGQIGDANFYFQNRKSRKHLRL 116
Db 66 MLEKFGAVGDANFYFQNRSSRRQR 94

RESULT 5
Q9FNM1 ID Q9FNM1 PRELIMINARY; PRT; 122 AA.
AC Q9FNM1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MCL19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kocani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006698; BAB08246.1; -.
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
SQ SEQUENCE 122 AA; 14044 MW; AAE9D0C6FD655E23 CRC64;

Query Match 13.6%; Score 238.5; DB 10; Length 122;
Best Local Similarity 45.4%; Pred. No. 6.9e-13;
Matches 54; Conservative 17; Mismatches 37; Indels 11; Gaps 4;

QY 2 SSSKNWPSMFKSPKNNHHH---QHE-IDTPSYMHYSCNLSFSSDRIPDPKPNW 57
Db 3 SSSKNK-----SNKDDHNDQGHNTIHTPSMHTFEISNISSPSPVVDPKPEWK 56

QY 58 PKPEQIRLESIFNSGTINPPREIQRIRIQEYQ-OIGDANFYFQNRKSRKHLRL 115
Db 57 PNQHQAILLELFGTGVNPSLTSIKQITIKLSYGEVDADYKWFHNKRSRKPRL 115

RESULT 6
Q9ZVF5 ID Q9ZVF5 PRELIMINARY; PRT; 295 AA.
AC Q9ZVF5;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

```
DE PUTATIVE HOMEODOMAIN TRANSCRIPTION FACTOR.
GN AT2G01500.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MEDLINE=20033487; Pubmed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DE EMBL; AC005560; AAC67326.1; -.
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
SQ SEQUENCE 295 AA; 34187 MW; D0B95CD2C9A08E5 CRC64;

Query Match 12.4%; Score 217.5; DB 10; Length 295;
Best Local Similarity 28.3%; Pred. No. 1.4e-10;
Matches 78; Conservative 32; Mismatches 91; Indels 75; Gaps 12;

QY 11 MFKSKPCNNHHQ-----HEIDTPSYMHYSCNLSFSSDRIPDPKP-RWNPKPE 61
Db 47 LFFSDINGNDHHQLITASSGEHDIDE-----RKNIPAAATLRNPTPE 91

QY 62 QIRLESIFNSGTINPPREIQRIRIQEYQIGDANFYFQNRKSRKHLRHVHKS 121
Db 92 QITTEELYRSGTRPTTTEIQIQAASKLRKYGRIEKGVNFYFQNHKARLKR-RREG 150

QY 122 PKMSKDKTVPSTADHCFGVNQETGLYPVQNNELVVTPEAGFLFPVHNDPSAAQSAF 181
Db 151 GAITKPKHDKVDKSSSGGHRV---DQTKLCP-----SFPHTRPQ----- 186

QY 182 GFGDFVPVVTTEEGMAFSTVNVGNVLE---TNENFDKIPAINLYGDDGGNGGCPPLTV 238
Db 187 -----PQHELDPASYNKDNANNEDHGTTEESQORASEVGYATWRN-----LV 230

QY 239 PLTINQSQEK--RDVGLSG-GEVDGD----NVYPVR 267
Db 231 TWSITQPEEINIDENVNGEETRDNRNLNLPVR 266

RESULT 7
Q9SB92 ID Q9SB92 PRELIMINARY; PRT; 291 AA.
AC Q9SB92;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE WUSCHEL PROTEIN.
GN WUSCHEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
```

Query Match	10.8%;	Score 188.5;	DB 10;	Length 350;
Best Local Similarity	31.0%;	Pred. No. 5e-08;		
Matches	44;	Conservative	26;	Mismatches 61; Indels 11; Gaps
QY	18	NNHHHGHQHTDTPSYMHYSCNLSLSSFFSGSDRIDPKP-----RWNPKEQIRILESIFNS 72		
Db	36	NTNSDRFNNAVVTMTAEQNKREMLMNSE--PQHPVVMVSSRWNTPTDGLRVLEELYRQ 93		
QY	73	GTINPPREETORIRIRLQEYGOIGDANVFYWFQNRKSRAKHLRVHHKSPKMSKKDKTVI 132		
Db	94	GTRTPSADHIQQTALRLRYKIEGRNVFYWFQNHKARERQKRRQME----TGHSETVL 149		
QY	133	PSTDADHCFGFGVNETGLYPVQ 154		
Db	150	STASLVSNHGFDKDKPPGYKVE 171		
RESULT	9			
Q9SL51		PRELIMINARY;	PRT;	293 AA.
AC	Q9SL51;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	PUTATIVE HOMEODOMAIN TRANSCRIPTION FACTOR.			
GN	AT2G17950.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RC	MEDLINE=20083487; PubMed=10617197;			
RX	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.			
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.			
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,			
RA	Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,			
RA	Salzberg S.L., Fraser C.M., Venter J.C.;			
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RT	thaliana.";			
RL	Nature 402:761-768(1999).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RC	Lin X.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC006201; AAD20131.1; -			
DR	InterPro; IPR001356; Homeobox.			
DR	SMART; SM00389; HOX; 1.			
DR	PROSITE; PS50071; HOMEBOX_2; 1.			
KW	Homeobox; DNA-binding; Nuclear protein.			
SW	SEQUENCE 293 AA; 33150 MW; 01A4397B39775D6B CRC64;			
Query Match	10.7%;	Score 188;	DB 10;	Length 293;
Best Local Similarity	30.5%;	Pred. No. 4.4e-08;		
Matches	43;	Conservative	31;	Mismatches 61; Indels 6; Gaps
QY	15	KPCNNHHHGHQHTDTPSYMHYSCNLSLSSPSSSDRIDPKPRWNPKEQIRILESI-FNSG 73		
Db	2	EPHQHQHH--HQADQESGNNNNKSGSGGVTCTQ--TSTRWTPTTEQIKILKELYNNA 57		
QY	74	TINPPREETORIRIRLQEYGOIGDANVFYWFQNRKSRAKHLRVHHKSPKMSKKDKTVIP 133		
Db	58	IRSTAQDQIQITARLQKQFQIEGKNVYWFQNHKARERQKRENGTNGTWTTPSSNSVM 177		
QY	134	STDADHCFGFGVNETGLYPVQ 154		


```

Db 118 MAANDHYHPLHHHGV-PHQ 137
RESULT 10
Q9C775
ID Q9C775 PRELIMINARY; PRT; 165 AA.
AC Q9C775
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 19.2 KDA PROTEIN.
GN FilB9.18
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salancubut M., Lencke K., Rieger M., Ansoerge W., Unsel M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Cholsne N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Cascuberta E.,
RA Nonfort A., Argirion A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rued S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana.";
RL Nature 408:820-822(2000).
DR EMBL; AC073395; AAG50976.1; -.
DR InterPro: IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 165 AA; 19154 MW; 39B2AE354E0C6344 CRC64;

Query Match 10.1%; Score 177.5; DB 10; Length 165;
Best Local Similarity 35.9%; Pred. No. 1.6e-07;
Matches 37; Conservative 18; Mismatches 35; Indels 13; Gaps 1;

QY 55 RWNPKPEQIRILEISNGTINPREETQIRIRLOEYGOIGDANFYWFQNRKSRK 114
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 RWNPTVQLKILDLFRAGLRTTTDQIQKISTELFSGKIESKNFYWFQNRKARQK 83
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 115 LRVHHKSPKMSKDKTVPSTADHCFGFVNQETGLYPVQ 157
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 84 RHVFEISEDCQEEKVIETLQ-----LFPVNSFE 113

RESULT 11
Q9LM83
ID Q9LM83 PRELIMINARY; PRT; 197 AA.
AC Q9LM83
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE F2D10.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaverl A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F2D10 from chromosome
1.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC069251; AAF80617.1; -.
DR InterPro: IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 197 AA; 22400 MW; 9BC78D91CC88DE9F CRC64;

Query Match 10.0%; Score 175; DB 10; Length 197;
Best Local Similarity 34.9%; Pred. No. 3.4e-07;
Matches 38; Conservative 22; Mismatches 35; Indels 14; Gaps 3;

QY 53 KPRWNPKPEQIRILEISNGTINPREETQIRIRLOEYGOIGDANFYWFQNRKSRK 112
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 77 RPRWPTTQQLLENLYKEGSGTPNPRRIKEITMELSEHGQIMEKNVYHFWNRRARSK 136
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 113 HKLVRVHKSP-----KMSKKDKTVPSTD----ADHCFGFVNQETGLYP 152
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 RK-----QPPTTTTSSQADDAVTTTEERGRGCGDGGFESYEHILFP 180

RESULT 12
Q9FSR5
ID Q9FSR5 PRELIMINARY; PRT; 262 AA.
AC Q9FSR5
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE HOMEODOMAIN TRANSCRIPTION FACTOR.
GN H0212B02.16.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Zhou B., Feng Q., Huang Y.C., Chen Z.H., Li Y., Zhu J.J.,
RA Tang Y.S., Zhao Q., Liu Y.L., Mu J., Yu Z., Fan D.L., Chen L.,
RA Wang Q.J., Zhang L., Liu Y.Q., Yu S.L., Zhu J., Liu X.H., Hu X.,
RA Lei H.Y., Zhang Y.J., Wang R., Li C., Lu Y., Chen X.C., Zhang Y.,
RA Hu H., Jia P.X., Li T., Qian Y.M., Ying K., Hong G.F.;
RT "Oryza sativa indica (Guangluai4) genomic DNA, chromosome4, BAC
clone:H0212B02.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL442007; CAC09359.1; -.
DR InterPro: IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein.

```

OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA	Kim C., Altafi H., Bel O., Chin C., Chou J., Choi E., Conn L.,
RA	Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA	Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA	Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA	Toriuni M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA	Ecker J.R.;
RT	"Genomic sequence for Arabidopsis thaliana BAC F2D10 from chromosome
RT	I."
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AC069251; AAF80616.1; -.
DR	InterPro; IPR001356; Homeobox.
DR	Pfam; PF00046; homeobox; 1.
DR	SMART; SM00389; HOX; 1.
DR	PROSITE; PS50071; HOMEBOX_2; 1.
SQ	SEQUENCE 211 AA; 23976 MW; D8E19461EAF285D2 CRC64;
Query Match	9.8%; Score 171.5; DB 10; Length 211;
Best Local Similarity	45.3%; Pred. No. 7.3e-07;
Matches	34; Conservative 15; Mismatches 25; Indels 1; Gaps 1;
QY	40 LSSFSFSDRPPDKPRNPKEQRIRESINSGTINPREIQIRIRLOFYGOIGDAN 99 : : : : : : :
Db	81 MGASSSHRI-STHRWTPTSTQLLESIVDEGSGTPRRRIRREIATLSHGQITETN 139 : : : : : : :
QY	100 VFYWFQNRKSRKHK 114 : : :
Db	140 VYNWFQNRARSKRK 154 : : :
RESULT 15	
Q9ZRZ4	PRELIMINARY; PRT; 261 AA.
ID Q9ZRZ4	AC Q9ZRZ4;
AC Q9ZRZ4	01-MAY-1999 (TEMBLrel. 10, Created)
DT DT	01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE HB2	HOMEODOMAIN PROTEIN (FRAGMENT).
GN HB2.	
OS Populus tremula x Populus tremuloides.	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	
OC eurosid II; Malpighiales; Salicaceae; Populus.	
OX NCBI_TaxID=47664;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=CAMBIAL REGION;	
RC Hertzberg M., Olsson O.;	
RT "Molecular characterisation of a novel plant homeobox gene expressed	
RT in the maturing xylem zone of Populus tremula x tremuloides.";	
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; AJ010811; CAA09367.1; -.	
DR InterPro; IPR001356; Homeobox.	
DR Pfam; PF00046; homeobox; 1.	
DR SMART; SM00389; HOX; 1.	
DR PROSITE; PS50071; HOMEBOX_2; 1.	
KW Homeobox; DNA-binding; Nuclear protein.	
FT NON_TER 1	
SQ SEQUENCE 261 AA; 29439 MW; 767AF56899932C27D CRC64;	
Query Match	9.7%; Score 170; DB 10; Length 261;
Best Local Similarity	34.3%; Pred. No. 1.3e-06;
Matches	36; Conservative 19; Mismatches 34; Indels 16; Gaps 2;
QY	53 KPRNPNPEQRIRESIFNSGTINPPREIQIRIRLOEQYIGDANFYFWFNKRSAK 112 : : : : : : : : : : : :

Db 117 RQRWTPVQLILERIFDOGNTPSKOKIKEITSELQHQISETNVYNWQNRARSK 176
Oy 113 HK-----LRVHKSPKMSKKDKTVI-----PSTDADHCF 141
Db 177 RKQLVASSNNAESEVETEVDLSNEKKKPEIFHAQONPPRAEDLCF 221

Search completed: August 27, 2002, 20:29:11
Job time: 433 sec

THIS PAGE BLANK (USPTO)

PI	Olsson O, Hertzberg M;
XX	
DR	WPI; 1999-610855/52.
DR	P-PSDB; AAV32095.
XX	
PT	New isolated plant homeobox genes, used to develop products for
PT	regulating the fiber properties of fibrous plants, particularly woody
PT	plants -
XX	
CS	Claim 4; Page 30-31; 36pp; English.
XX	
CC	This is the nucleotide sequence of cDNA for the novel homeobox gene
CC	PtHbl1 of hybrid aspen (<i>Populus tremula</i> x <i>Populus tremuloidea</i>). The
CC	cDNA was isolated from a cambial cDNA library of the hybrid aspen.
CC	The present invention discloses a novel class of homeobox genes,
CC	characterized by PtHbl1 and PtHbl2 (see AA220289), that influence cell
CC	differentiation and growth of fibrous plants. The PtHbl1 gene
CC	displays a tissue-specific expression, being active in the xylem
CC	maturization zone of the cambial region. The PtHbl2 gene is active
CC	in the earlier developmental phases on both sides of the cambium,
CC	as well as in the cambium itself. The isolated genes (and
CC	antisense sequences) can be used in claimed methods for the
CC	regulation of the fibre properties of fibrous plants.
CC	Transgenic fibrous plants are claimed that can be conifer or
CC	dicotyledonous softwood trees or annual angiosperms. The
CC	properties of fibrous raw material can be modified already in the
CC	growing plant through the functional inclusion of the homeobox gene.
XX	
SQ	Sequence 1136 BP; 387 A; 197 C; 280 G; 271 T; 1 other;
	Query Match 4.9%; Score 59.6; DB 20; Length 1136;
	Best Local Similarity 56.7%; Pred. No. 0.013;
	Matches 110; Conservative 0; Mismatches 84; Indels 0; Gaps
QY	185 agatcctaaccgcagatgaattccctaaacggagcagattaggatactcgataacttt 244
Dd	
Db	355 aggttcgaggcagcggtggtcaccaacgccagcgaacctccaactccttgtgacgat 414
QY	245 caattccggtactattaaccacacctagagagagagtccaagaatccggtccggcttca 304
Dd	
Db	415 cgacaattgcaatgcgactccaggcaggcagaagaatcaaagatataacccgggaactgc 474
QY	305 agaatatggtcaaatcgdtgcagcaaacgctgttttactggtttccaaacgggaactgc 364
Dd	
Db	475 acacaatgccc aaattctgaaacaaaatgtctcaattggttcacaaacaggagactgc 534
QY	365 agcaaaacacaaagc 378
Dd	
Db	535 tccaagaagaagc 548

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 20:14:52 ; Search time 28.12 Seconds
(without alignments)
1283.748 Million cell updates/sec

Title: US-09-787-737-2

Perfect score: 1751

Sequence: 1 MSSSKNWPMSFKSRPCNN.....TDEFGVTYQLQNGAIYYLI 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1751	100.0	325	22	Arabidopsis thalia
2	645.5	36.9	378	22	Arabidopsis thalia
3	325.5	18.6	253	22	Corn Wuschel (WUS)
4	212.5	12.1	271	21	Arabidopsis thalia
5	200	11.4	292	21	Arabidopsis thalia
6	188	10.7	123	21	Pinus radiata tran
7	187	10.7	123	21	Corn Wuschel (WUS)
8	187	10.7	220	22	Corn Wuschel (WUS)
9	175.5	10.0	217	22	Corn Wuschel (WUS)
10	173	9.9	221	22	Soybean WUS protei
11	172.5	9.9	231	22	Corn Wuschel (WUS)

12	170	9.7	134	22	AAU01528	Soybean WUS protei
13	170	9.7	261	20	AAU01528	Aspen homeobox pro
14	167.5	9.6	217	20	AAU01528	Aspen homeobox pro
15	167.5	9.6	238	22	AAU01523	Corn Wuschel (WUS)
16	167	9.5	236	21	AAU01523	Arabidopsis thalia
17	167	9.5	267	21	AAU01523	Arabidopsis thalia
18	167	9.5	268	21	AAU01524	Arabidopsis thalia
19	166	9.5	237	22	AAU01524	Corn Wuschel (WUS)
20	140.5	8.0	221	22	AAU01527	Corn Wuschel (WUS)
21	135.5	7.7	89	22	AAU01526	Corn Wuschel (WUS)
22	129	7.4	146	21	AAU01526	Eucalyptus grandis
23	115.5	6.6	101	21	AAU01526	Eucalyptus grandis
24	109	6.2	77	22	AAU01521	Corn Wuschel prote
25	101	5.8	129	21	AAU01521	Eucalyptus grandis
26	101	5.8	179	21	AAU01521	Eucalyptus grandis
27	100	5.7	272	22	AAU01521	Arabidopsis thalia
28	99.5	5.7	689	19	AAU01521	H. pylori GHO 992
29	99.5	5.7	689	22	AAU01521	Helicobacter pylor
30	98.5	5.6	958	18	AAU01521	Yeast alanyl-tRNA
31	98	5.6	255	21	AAU01521	Arabidopsis thalia
32	95	5.4	673	19	AAU01521	S. pneumoniae deri
33	93.5	5.3	848	22	AAU01521	Novel human diagno
34	93.5	5.3	1482	22	AAU01521	Novel human diagno
35	93	5.3	1772	22	AAU01521	Novel human diagno
36	92.5	5.3	630	19	AAU01521	Plasmodium falcipa
37	92	5.3	508	21	AAU01521	Human hepatocyte n
38	91	5.2	359	21	AAU01521	Human nucleic acid
39	91	5.2	420	21	AAU01521	Human secreted pro
40	91	5.2	500	21	AAU01521	Arabidopsis thalia
41	91	5.2	575	21	AAU01521	Arabidopsis thalia
42	91	5.2	598	21	AAU01521	Arabidopsis thalia
43	91	5.2	631	19	AAU01521	Human hepatocyte n
44	91	5.2	631	19	AAU01521	Human native hepat
45	91	5.2	632	22	AAU01521	Drosophila melanog

ALIGNMENTS

RESULT 1
AAB60642
ID AAB60642 standard; Protein; 325 AA.
XX
AC AAB60642;
XX
DT 04-MAY-2001 (first entry)
XX
DE Arabidopsis thaliana M6 homeodomain protein.
XX
KW Homeodomain-like; homeobox gene; M6; plant differentiation;
KW MSH; many shoot; expression control; adventitious bud formation;
KW branching induction; cytokinin independent; transgenic plant; crop;
KW agriculture.
XX
OS Arabidopsis thaliana.
XX
PN WO200107618-A1.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-JP04904.
XX
PR 22-JUL-1999; 99JP-0207995.
XX
PA (SUNR) SUNTORY LTD.
XX
PI Kakimoto T;
XX
DR WPI; 2001-182796/18.
XX
DR N-PSDB; AAF59786.
XX
PT Arabidopsis thaliana originated homeobox genes encoding proteins
PT participating in differentiation with ability to promote adventitious

PT bud formation and branching induction, applicable in improving plant
 PT cells and plants e.g. for crops -
 PS Claim 1; Page 23-24; 34pp; Japanese.
 XX
 XX The invention relates to two novel Arabidopsis thaliana proteins M6
 CC and M8 (AAB60642 and AAB60643, respectively), and to the cDNAs
 CC encoding them (AAF60642 and AAF60643). The proteins contain homeodomain-
 CC like sequences and participate in differentiation, having the ability
 CC to promote adventitious bud formation and branching induction by
 CC controlling expression of the many shoot (MSH) gene without the
 CC requirement for cytokinin. The invention also relates to expression
 CC vectors and host cells which contain M6 or M8 nucleic acid sequences,
 CC the recombinant production of the M6 or M8 protein, a plant or plant
 CC cell transfected with the M6 or M8 gene, and methods for inducing
 CC differentiation, adventitious bud formation and/or plant branching via
 CC expression of the transfected M6 or M8 gene. Nucleic acids encoding the
 CC homeodomain proteins may be used to generate transgenic plants with
 CC improved adventitious bud formation and branching for use e.g., as
 CC agricultural crops. The present sequence represents the Arabidopsis
 CC thaliana M6 protein.
 XX Sequence 325 AA;
 SQ
 Query Match 100.0%; Score 1751; DB 22; Length 325;
 Best Local Similarity 100.0%; Pred. No. 2e-163;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSSNNKWPMSFKPCNNHHQHHEIDTPSYMHYSCNLSLSSFSDDRIDPKPRWNPKP 60
 Db 1 massnknwpsmfkskpcnnhhqhheidtpsymhyscnlsssfssddridpkprwnpkp 60
 QY 61 EOTRILESFNSGTINPPREETQIRIRLOEQIGDANVFYFQNRKSPRAKHLRVHKK 120
 Db 61 eqirilesfnsgtlnppreetiqrirleqyvgigdauvfyrqnrksrakhkrlrvhkk 120
 QY 121 SPKMSKKDKTIVIPSTDADHCFGVNQTGLYPQNNELVTEPAGFLFPVNDPSPAQA 180
 Db 121 spkmskkdkttvtpstdadhdhcfgvnqtglypvqnnelvtepagflfpvndpsaaqa 180
 QY 181 FPGDFVPVPTVEGMAFSTVNGVNLSTNENFDKIPAINLYGGDNGGNCFFPPTVPL 240
 Db 181 fgdfgvvpvptveegmafstvcnngvnlstnenfdkipainlyggdnggncffpptvpl 240
 QY 241 TINQSEKRDVGLSGGEDYGDNVYVPMVTFINEMPIEVVSGLFNFKAAFGNDVILNSF 300
 Db 241 tinqsekrdvglsggedygdnvypvmtvfinempienvvsglfnfkaafgndavilnsf 300
 QY 301 GQPILTDEFQVTVQPLQNGAIYYLI 325
 Db 301 gqpiltdefgvtvqplngaiyyli 325
 RESULT 2
 AAB60643
 ID AAB60643 standard; Protein; 378 AA.
 XX
 AC AAB60643;
 XX
 DT 04-MAY-2001 (first entry)
 XX
 DE Arabidopsis thaliana M8 homeodomain protein.
 XX
 KW Homeodomain-like; homeobox gene; M8; plant differentiation;
 KW MSH; many shoot; expression control; adventitious bud formation;
 KW branching induction; cytokinin independent; transgenic plant; crop;
 KW agriculture.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200107618-A1.

PD 01-FEB-2001.
 XX
 PF 21-JUL-2000; 2000WO-JP04904.
 XX
 PR 22-JUL-1999; 99JP-0207995.
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Kakimoto T;
 XX
 DR WPI: 2001-182796/18.
 DR N-PSDB; AAF59787.
 XX
 XX Arabidopsis thaliana originated homeobox genes encoding proteins
 PT participating in differentiation with ability to promote adventitious
 PT bud formation and branching induction, applicable in improving plant
 PT cells and plants e.g. for crops -
 XX
 XX Claim 4; Page 28-30; 34pp; Japanese.
 PS
 XX The invention relates to two novel Arabidopsis thaliana proteins M6
 CC and M8 (AAB60642 and AAB60643, respectively), and to the cDNAs
 CC encoding them (AAF60642 and AAF60643). The proteins contain homeodomain-
 CC like sequences and participate in differentiation, having the ability
 CC to promote adventitious bud formation and branching induction by
 CC controlling expression of the many shoot (MSH) gene without the
 CC requirement for cytokinin. The invention also relates to expression
 CC vectors and host cells which contain M6 or M8 nucleic acid sequences,
 CC the recombinant production of the M6 or M8 protein, a plant or plant
 CC cell transfected with the M6 or M8 gene, and methods for inducing
 CC differentiation, adventitious bud formation and/or plant branching via
 CC expression of the transfected M6 or M8 gene. Nucleic acids encoding the
 CC homeodomain proteins may be used to generate transgenic plants with
 CC improved adventitious bud formation and branching for use e.g., as
 CC agricultural crops. The present sequence represents the Arabidopsis
 CC thaliana M8 protein.
 XX
 SQ Sequence 378 AA;
 Query Match 36.9%; Score 645.5; DB 22; Length 378;
 Best Local Similarity 40.4%; Pred. No. 9.3e-55;
 Matches 166; Conservative 42; Mismatches 84; Indels 119; Gaps 17;
 QY 1 MSSNNKWPMSFKPCNNHHQHHEIDTPSYMHYSCNLSLSSFSDDRIDPKPRW 56
 Db 1 massnknwpsmfkskpcnnhhqhheidtpsymhyscnlsssfssddridpkprw 56
 QY 57 NPKPEQIRILESFNSGTINPPREETQIRIRLOEQIGDANVFYFQNRKSPRAKHLR 116
 Db 57 npkpeqirileafnsgmvpnpreeirirleqyvgigdauvfyrqnrksrkhkrl 116
 QY 117 VHH-----KSPKMSKKDKTIVIP-----STDADHCFGVNQTG 149
 Db 117 ilhnhskhslpqtqppqpqpssssssssskskprksknknntnlslg-gsqmng 175
 QY 150 LYPQNNELVTEPAGFLFPVHN-----DPSAAQSAFAGF--GDFV--VPVVTBEGMAFST 200
 Db 150 mfp-----pepa-flfpvstvvgfegitvssqglfsgdmieqkpkaptctglllse 226
 QY 201 VNNG-VNLETNE-----NFDKIPAINLYGGDNGGNGG 231
 Db 227 imngsvsygthhqqhisekeveemrkmlqqptqtcicayattnhqiasyn-----nnnnnn 282
 QY 232 C----FPP-----LTVPLTINQSEKRDVGLSGGEDVGDNVYVPMVTFINE 274
 Db 283 imhippttstatittshslatvpsdqlqvad-----arirvine 327
 QY 275 MPIEVVSGLFNFKAAFGNDVILNSFQPILTDEFQVTVQPLQNGAIYYLI 325
 Db 328 melvssgpfvndafgeevvllinsagpivtdeygvallhplqhgsayyli 378


```
RESULT 3
AAU01525 ID AAU01525 standard; Protein; 253 AA.
XX
AC AAU01525;
XX
DT 16-JUL-2001 (first entry)
XX
DE Corn Wuschel (WUS) protein from clone p0058_chpab57r.
XX
KW Wuschel; WUS; apical meristem; organ formation; homeodomain protein;
KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
KW plant transformation; transgenic plant; p0058_chpab57r.
XX
OS Zea mays.
XX
PN WO200123575-A2.
XX
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26648.
XX
PR 30-SEP-1999; 99US-0157216.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Lowe K;
XX
DR WPI; 2001-258137/26.
DR N-PSDB; AAS02222.
XX
XX
Novel nucleic acid fragments encoding WUS proteins useful for
transiently modulating WUS protein level in plant cells, as probes for
genetically and physically mapping WUS genes and as markers
Claim 3; Fig 1; 6lpp; English.
XX
The sequence represents a Wuschel (WUS) protein, encoded by cDNA from a
clone. The WUS protein plays a key role in initiation and maintenance of
the apical meristem of a plant which contains stem cells, allowing for
organ formation. WUS encodes a homeodomain protein which is thought to
function as a transcriptional regulator. The DNA or its complement is
useful for transiently modulating the level of WUS protein in a plant
cell and meristem proliferation can be induced through its addition. This
involves transforming a plant with WUS DNA to produce a transformed
meristem which is grown into a regenerated plant (corn, soybean, wheat,
rice, alfalfa, sunflower, canola or cotton). The protein and DNA are
useful in the identification of sequences that affect WUS level or enzyme
activity, or that encode homologous proteins from the same or other plant
species. The polypeptides can then be used to prepare antibodies.
Fragments of the DNA may be used as probes to map the genes that they are
part of and as markers for traits linked to those genes, which is useful
in plant breeding to develop lines with desired phenotypes. The nucleic
acid sequences are useful to create transgenic plants in which the WUS
polypeptides are present in higher or lower levels than normal or in cell
types or developmental stages in which they are not normally found.
XX
Sequence 253 AA;
Query Match 18.6%; Score 325.5; DB 22; Length 253;
Best Local Similarity 31.3%; Pred. No. 1.4e-23;
Matches 92; Conservative 40; Mismatches 109; Indels 53; Gaps 7;
QY 40 LSSSFSSDRIPDKP---RWNPKPQRIRESIFNSGTTNPPRETIQRIRLQYQGIG 96
DB 1 meggisperhaaeprvrtwtpkpeqillilesifnsgmvpnpkdetvirkilerfgvg 60
QY 97 DANVFVWFQNRKSRKAKHLRVRHHSFKMSKDKTVPSTDADHCFCGVNQETGLXPQNN 156
DB 61 danvfvyfqnrsvrrrrrqrqlqaaqaassssssgspp-----sglap---g 104
```

```
QY 157 ELVVTPEAGFLFPVHNDPSSAAQSAFGFGDFVVPVVTTEGMAFSTVANGVNLETNENFDKI 216
DB 105 hatasastagmfahgatygsasasw-----ppppscgmmgaldlygg-----gddi 150
QY 217 PAINLYGGDGGNGGNCFFPLTVPLTINQSOEKRDVGLSGGEDVGDNVY-----PVRMTVF 271
DB 151 falstrqmgvasgsggsassaavaahheqqq-----lyspcqpasmtvf 195
QY 272 INEMPIEVVYGLFNKAAFGNDVNLNSFGQPLITLDEFGVTVYQPLONGATYYLI 325
DB 196 lngvatevprgpidlrsmfgqdvmlvnhstagilpvneygvltsglmgesyflv 249
RESULT 4
AAG19619 ID AAG19619 standard; Protein; 271 AA.
XX
AC AAG19619;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21489.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
```

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0143624.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 22-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 23-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 212.5; DB 21; Length 271;
Best Local Similarity 28.3%; Pred. No. 1.9e-12;
Matches 76; Conservative 31; Mismatches 87; Indels 75; Gaps 12;

Qy	18	NNHHHQ-----HEIDTPSWHYSCNLSSTSSSSDRIPDPKP-RWNPKPEQIRILES	68
		: :	
Db	30	ngndhqlitassgehdide-----rknnpaaatlwnptpeqittlee	74
		: :	
Qy	69	IFNSGTINPPREIQIRIRLQEQYGOIGDANVFYWFQNBKSRAKHLRVHHKSPKWSKD	128
		: : : : : : :	
Db	75	lyrsgrtptedlqlasrlkyrgiekgnvfwtqnhkarelrkr-rreggaikph	133
		:	
Qy	129	KTVIPSTDADHCFFYQNQETGLYPVQNNELVTEPAGFLFPVHNDPSSAAQSAFGGDFV	188
		: :	
Db	134	kdvkdssgghrv---dqtklcp-----sfphtnrrpq-----	162
		: :	
Qy	189	PVTEEGMAFSTVNVNGVNL-----TNENFDKIPAINLYGGDGGNGCNPPLTVPLTINOS	245
		: : : : :	
Db	163	pqhdelpasynkdnannedhgttesdqravegkyatwn-----lvtwsiltqq	213
		: :	
Qy	246	QEK-RDVGLSG-GEVDGD-----NVYPVR	267
		: : : :	
Db	214	peelnidenvgeeeetrdrtnlilpvr	242
		: : : :	
RESULT	5		
AAG42056			
ID	AAG42056	standard; Protein; 292 AA.	
XX	AAG42056;		
XX			
DT			
XX	18-OCT-2000	(first entry)	
XX			
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 52402.	
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promotio		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0123788.	
PR	23-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0128645.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	

CC Eucalyptus grandis or Pinus radiata. The present sequence is one such
CC transcription factor. The transcription factor may be used to produce a
CC plant having modified gene expression such as a woody plant e.g. a
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
CC to modify the activity of a polypeptide in a plant. The transcription
CC factors of the present invention are members from the following families
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
CC zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
CC Cys2His2, CCAAT box elements and MYB.

XX Sequence 123 AA;

Query Match 10.7%; Score 188; DB 21; Length 123;
Best Local Similarity 43.9%; Pred. No. 1.6e-10;
Matches 36; Conservative 14; Mismatches 28; Indels 4; Gaps 1;

Qy 55 RWNPKPEQIRILEIFNSGTINPREETQIRIRLQEVGIGDANVFYWFQNRKSRKHK 114

|||||:|||||||:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

Db 14 rwnptpdqirilemfykgmtptnaeglehitagirygklegknvfywfgnhkarerqk 73

Qy 115 LR----VHHKSPKMSKDKTKVI 132

: | : : || | |

Db 74 qkrnssmhqvaataakktptt1 95

RESULT 7

ID AAU01530 standard; Protein; 212 AA.

XX AAU01530;

16-JUL-2001 (first entry)

XX Corn Wuschel (WUS) protein from clone ses4d_pk0033_c8.

XX Wuschel; WUS; apical meristem; organ formation; homeodomain protein;
KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
KW plant transformation; transgenic plant; ses4d_pk0033_c8.

XX Glycine max.

XX WO200123575-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US266648.

XX 30-SEP-1999; 99US-0157216.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Cahoon RE, Lowe K;

XX WPI; 2001-258137/26.

XX N-PSDB; AAS02227.

XX Novel nucleic acid fragments encoding WUS proteins useful for
PT transiently modulating WUS protein level in plant cells, as probes for
PT genetically and physically mapping WUS genes and as markers -

XX Claim 3; Fig 1; 61pp; English.

XX The sequence represents a Wuschel (WUS) protein, encoded by cDNA from a
CC clone. The WUS protein plays a key role in initiation and maintenance of
CC the apical meristem of a plant which contains stem cells, allowing for
CC organ formation. WUS encodes a homeodomain protein which is thought to
CC function as a transcriptional regulator. The DNA or its complement is
CC useful for transiently modulating the level of WUS protein in a plant
CC cell and meristem proliferation can be induced through its addition. This
CC involves transforming a plant with WUS DNA to produce a transformed

CC meristem which is grown into a regenerated plant (corn, soybean, wheat,
CC rice, alfalfa, sunflower, canola or cotton). The protein and DNA are
CC useful in the identification of sequences that affect WUS level or enzyme
CC activity, or that encode homologous proteins from the same or other plant
CC species. The polypeptides can then be used to prepare antibodies. They are
CC fragments of the DNA may be used as probes to map the genes that they are
CC part of and as markers for traits linked to those genes, which is useful
CC in plant breeding to develop lines with desired phenotypes. The nucleic
CC acid sequences are useful to create transgenic plants in which the WUS
CC polypeptides are present in higher or lower levels than normal or in cell
CC types or developmental stages in which they are not normally found.

XX Sequence 212 AA;

Query Match 10.7%; Score 187; DB 22; Length 212;
Best Local Similarity 29.8%; Pred. No. 4.4e-10;
Matches 62; Conservative 19; Mismatches 55; Indels 72; Gaps 9;

Qy 43 SFSSD-----RIPDPKPRWNPKEQIRILEIFNSGTINPREETQIRIRLQEVGIG 96

|||||:|||||||:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

Db 3 shssdaeaenvrthssvswptkeqldmlelnlykgirtpteqiqqitsrlrayghle 62

Qy 97 DANVFYWFQNRKSRKHKLRVHHKSPKMSKDKTKVIPST---DADH-----CFGFVNQ 146

|||||||:| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

Db 63 gknvfywfgnhkarqrkl-----mkqgtiaysnrflrashpicnvcapycicq 112

Qy 147 ETG--LYPVQNNELV-----VTEPAGF-----LFPV----- 170

: | | | | :| | | | :| | | | :| | | | :| | | | :| | | |

Db 113 rsgfsfyqqskvlasgsgisltgpiqmrgmfdgmssehpdcnrevltfiphtgtlke 172

Qy 171 ---HNDPSSAAQSAFGFDFVPVPTVEEG 195

| | | | | :| | | | :| | | | :| | | | :| | | |

Db 173 kthqvpslasts-----vvavdedg 193

RESULT 8

AAU01522

ID AAU01522 standard; Protein; 220 AA.

XX AAU01522;

16-JUL-2001 (first entry)

XX Corn Wuschel (WUS) protein from clone cpilc_pk012_p19.

XX Wuschel; WUS; apical meristem; organ formation; homeodomain protein;
KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
KW plant transformation; transgenic plant; cpilc_pk012_p19.

XX Zea mays.

XX WO200123575-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US266648.

XX 30-SEP-1999; 99US-0157216.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Cahoon RE, Lowe K;

XX WPI; 2001-258137/26.

XX N-PSDB; AAS02219.

XX Novel nucleic acid fragments encoding WUS proteins useful for
PT transiently modulating WUS protein level in plant cells, as probes for
PT genetically and physically mapping WUS genes and as markers -
XX Claim 3; Fig 1; 61pp; English.

XX The sequence represents a Wuschel (WUS) protein, encoded by cDNA from a
 CC clone. The WUS protein plays a key role in initiation and maintenance of
 CC the apical meristem of a plant which contains stem cells, allowing for
 CC organ formation. WUS encodes a homeodomain protein which is thought to
 CC function as a transcriptional regulator. The DNA or its complement is
 CC useful for transiently modulating the level of WUS protein in a plant
 CC cell and meristem proliferation can be induced through its addition. This
 CC involves transforming a plant with WUS DNA to produce a transformed
 CC meristem which is grown into a regenerated plant (corn, soybean, wheat,
 CC rice, alfalfa, sunflower, canola or cotton). The protein and DNA are
 CC useful in the identification of sequences that affect WUS level or enzyme
 CC activity, or that encode homologous proteins from the same or other plant
 CC species. The polypeptides can then be used to prepare antibodies.
 CC Fragments of the DNA may be used as probes to map the genes that they are
 CC part of and as markers for traits linked to those genes, which is useful
 CC in plant breeding to develop lines with desired phenotypes. The nucleic
 CC acid sequences are useful to create transgenic plants in which the WUS
 CC polypeptides are present in higher or lower levels than normal or in cell
 CC types or developmental stages in which they are not normally found.
 XX
 SQ Sequence 220 AA;

Query Match 10.7%; Score 187; DB 22; Length 220;
 Best Local Similarity 38.9%; Pred. No. 4.6e-10;
 Matches 37; Conservative 14; Mismatches 36; Indels 8; Gaps 2;
 QY 55 RWNPKPEQIRILESIENSGTINPPREIORIRLOBYGOIGDANVYFQNRKSRKHK 114
 DB 14 rwnptaeqkvvlcelfraglrtpsteqqrtrishlsafgkvesknvfwfqnhkarerh- 72
 QY 115 LRVHHSKPKMSKDKTVIPSTADHCFGVNQETG 149
 DB 73 ---hhkrrirgass-----spdsgrgsnneedg 100

RESULT 9
 AAU01529
 ID AAU01529 standard; Protein: 217 AA.
 AC AAU01529;
 DT 16-JUL-2001 (first entry)
 XX Corn Wuschel (WUS) protein from clone scrlc_pk001_d2.
 DE
 KW Wuschel; WUS; apical meristem; organ formation; homeodomain protein;
 KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
 KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
 KW plant transformation; transgenic plant; scrlc_pk001_d2.
 XX
 OS Glycine max.
 XX
 XX WQ200123575-A2.
 XX
 XX 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-US26648.
 XX
 XX 30-SEP-1999; 99US-0157216.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Cahoon RE, Lowe K;
 XX
 XX WPI; 2001-258137/26.
 XX
 XX N-PSDB; AAS02226.
 XX
 XX Novel nucleic acid fragments encoding WUS proteins useful for
 XX transiently modulating WUS protein level in plant cells, as probes for
 XX genetically and physically mapping WUS genes and as markers -
 XX

PS Claim 3; Page 55-56; 6lpp; English.
 XX
 CC The sequence represents a Wuschel (WUS) protein, encoded by cDNA from a
 CC clone. The WUS protein plays a key role in initiation and maintenance of
 CC the apical meristem of a plant which contains stem cells, allowing for
 CC organ formation. WUS encodes a homeodomain protein which is thought to
 CC function as a transcriptional regulator. The DNA or its complement is
 CC useful for transiently modulating the level of WUS protein in a plant
 CC cell and meristem proliferation can be induced through its addition. This
 CC involves transforming a plant with WUS DNA to produce a transformed
 CC meristem which is grown into a regenerated plant (corn, soybean, wheat,
 CC rice, alfalfa, sunflower, canola or cotton). The protein and DNA are
 CC useful in the identification of sequences that affect WUS level or enzyme
 CC activity, or that encode homologous proteins from the same or other plant
 CC species. The polypeptides can then be used to prepare antibodies.
 CC Fragments of the DNA may be used as probes to map the genes that they are
 CC part of and as markers for traits linked to those genes, which is useful
 CC in plant breeding to develop lines with desired phenotypes. The nucleic
 CC acid sequences are useful to create transgenic plants in which the WUS
 CC polypeptides are present in higher or lower levels than normal or in cell
 CC types or developmental stages in which they are not normally found.
 XX
 SQ Sequence 217 AA;

Query Match 10.0%; Score 175.5; DB 22; Length 217;
 Best Local Similarity 35.2%; Pred. No. 6.1e-09;
 Matches 43; Conservative 20; Mismatches 34; Indels 25; Gaps 5;
 QY 55 RWNPKPEQIRILESIENSGTINPPREIORIRLOBYGOIGDANVYFQNRKSRKHK 114
 DB 23 rwsptkeqldmlenfykqgirtsteqiqitrsrayglyegknvfwfqnhkarqrqk 82
 QY 115 LRVHHSKPKMSKDKTVIPSTADHCFGVNQETGLYPVQNNELVVTPE-----AGFLFP 169
 DB 83 lk-----qkqgsia-----yencflhas---hpicqn--vvcapyclqkgsfsy 122
 QY 170 VH 171
 DB 123 ph 124

RESULT 10
 AAU01532
 ID AAU01532 standard; Protein: 221 AA.
 AC AAU01532;
 XX
 XX 16-JUL-2001 (first entry)
 DT
 XX Soybean WUS protein from contig of ssm_pk0060_h4.
 DE
 KW Wuschel; WUS; apical meristem; organ formation; homeodomain protein;
 KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
 KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
 KW plant transformation; transgenic plant; ssm_pk0060_h4;
 KW NCBI GI No.4395781.
 XX
 XX Glycine max.
 XX
 XX WQ200123575-A2.
 XX
 XX 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-US26648.
 XX
 XX 30-SEP-1999; 99US-0157216.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Cahoon RE, Lowe K;
 XX
 XX WPI; 2001-258137/26.
 XX
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: August 27, 2002, 20:29:13 ; Search time 198.4 Seconds
(without alignments)
10505.716 Million cell updates/sec

Title: US-09-787-737-1

Perfect score: 1214

Sequence: 1 cttgactctcgattatcat.....ttaaaaaaaaaaaaaaaaaa 1214

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	100.0	1214	22	AAF59786
2	155.2	12.8	1518	22	AAF59787
3	69.8	5.7	1367	22	AAS02222
4	64.4	5.3	1326	21	AAC38580
5	63.8	5.3	371	21	AAC56282
6	60.8	5.0	506	22	AAS02225
7	60.8	5.0	844	22	AAS02226
C 8	60.8	5.0	6392	24	ABL32684
C 9	60.8	5.0	6392	24	ABL34506

10	59.6	4.9	1136	20	AAZ20288	Aspen homeobox gen
11	59.4	4.9	945	21	AAC42316	Arabidopsis thalia
12	57.8	4.8	741	22	AAS02227	Soybean Wuschel (W
13	55.8	4.6	4197	21	AAA70227	Plasmodium falcipa
C 14	55.8	4.6	6509	24	AAS61085	Human gene regulat
C 15	55.4	4.6	15548	24	ABL34155	Human immune syste
16	55.2	4.5	904	22	AAS02229	Soybean WUS cDNA f
C 17	55.2	4.5	6071	24	ABL32325	Human immune syste
C 18	55.2	4.5	6071	24	AAS61076	Human gene regulat
C 19	55	4.5	1190	20	AAZ20289	Human homeobox gen
C 20	55	4.5	7380	22	AAS45361	Chemically pretrea
C 21	54.2	4.5	7346	24	ABL32345	Human immune syste
C 22	54	4.4	1784	20	AAZ95521	Nucleic acid seque
C 23	53.6	4.4	1145	22	AAS02426	Human secreted pro
C 24	53.6	4.4	8467	24	ABL32109	Human immune syste
C 25	53	4.4	113515	24	ABL34175	Human immune syste
C 26	52.8	4.3	422	22	AAI15924	Human breast cance
27	52.6	4.3	6031	22	AAS46621	Tumour suppressor
C 28	52.2	4.3	7814	22	AAS46529	Tumour suppressor
C 29	51.8	4.3	310	22	AAH71505	Human cervical can
C 30	51.8	4.3	6101	24	ABL33474	Human immune syste
C 31	51.6	4.3	268	22	AAH70080	Human cervical can
C 32	51.6	4.3	7928	24	ABL32094	Human immune syste
C 33	51.6	4.3	7928	24	AAS61055	Human gene regulat
C 34	51.2	4.2	13449	24	ABL33385	Human immune syste
C 35	51	4.2	8210	24	AAS61282	Human gene regulat
C 36	50.8	4.2	5944	24	ABL32252	Human immune syste
C 37	50.8	4.2	6365	24	ABL32124	Human immune syste
C 38	50.8	4.2	15698	24	ABL34141	Human immune syste
39	50.8	4.2	17137	24	ABL32191	Human immune syste
40	50.6	4.2	6341	24	ABL33644	Human immune syste
41	50.4	4.2	648	21	AAC55917	Eucalyptus grandis
42	50.4	4.2	5487	24	ABL33598	Human immune syste
C 43	50.4	4.2	6109	24	ABL32326	Human immune syste
C 44	50.4	4.2	6109	24	AAS61077	Human gene regulat
C 45	50.4	4.2	7351	24	ABL32028	Human immune syste

ALIGNMENTS

RESULT 1

AAF59786

ID AAF59786 standard; cDNA; 1214 BP.

XX AAF59786;

AC AAF59786;

XX 04-MAY-2001 (first entry)

DT Arabidopsis thaliana M6 homeodomain protein cDNA.

DE Homeodomain-like; homeobox gene; M6; plant differentiation;

KW MSH; many shoot; expression control; adventitious bud formation;

KW branching induction; cytokinin independent; transgenic plant; crop;

KW agriculture; ss.

OS Arabidopsis thaliana.

XX WO200107618-A1.

PN 01-FEB-2001.

PF 21-JUL-2000; 2000WO-JP04904.

PR 22-JUL-1999; 99JP-0207995.

PA (SUNR) SUNTORY LTD.

PI Kakimoto T;

XX WPI; 2001-182796/18.

DR P-PSDB; AAB60642.

XX

PT Arabidopsis thaliana originated homeobox genes encoding proteins
PT participating in differentiation with ability to promote adventitious
PT bud formation and branching induction, applicable in improving plant
PT cells and plants e.g. for crops -
XX
PS Claim 3; Page 20-23; 34pp; Japanese.
XX
CC The invention relates to two novel Arabidopsis thaliana proteins M6
CC and M8 (AAB60642 and AAB60643, respectively), and to the cDNAs
CC encoding them (AAF60642 and AAF60643). The proteins contain homeodomain-
CC like sequences and participate in differentiation, having the ability
CC to promote adventitious bud formation and branching induction by
CC controlling expression of the many shoot (MSH) gene without the
CC requirement for cytokinin. The invention also relates to expression
CC vectors and host cells which contain M6 or M8 nucleic acid sequences,
CC the recombinant production of the M6 or M8 protein, a plant or plant
CC cell transfected with the M6 or M8 gene, and methods for inducing
CC differentiation, adventitious bud formation and/or plant branching via
CC expression of the transfected M6 or M8 gene. Nucleic acids encoding the
CC homeodomain proteins may be used to generate transgenic plants with
CC improved adventitious bud formation and branching for use e.g., as
CC agricultural crops. The present sequence represents cDNA encoding the
CC Arabidopsis thaliana M6 protein.
XX
SQ Sequence 1214 BP; 395 A; 228 C; 226 G; 365 T; 0 other;

Query Match 100.0%; Score 1214; DB 22; Length 1214;
Best Local Similarity 100.0%; Pred. No. 1.8e-223;
Matches 1214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctttagctcgtattatcatcattacacatcatcatgtctctcctcaaaacaaattggc 60
|||||
Db 1 ctttagctcgtattatcatcattacacatcatcatgtctctcctcaaaacaaattggc 60
QY 61 caagcatgttcaaatccaaaccttgaaacaataatcatcatcatcaacatgaatcgata 120
|||||
Db 61 caagcatgttcaaatccaaaccttgaaacaataatcatcatcatcaacatgaatcgata 120
QY 121 ctccattctacatgcactactctaaattgcaacctatcatcttcttctcgcagatcgga 180
|||||
Db 121 ctccattctacatgcactactctaaattgcaacctatcatcttcttctcgcagatcgga 180
QY 181 taccagatctcaaacggagatgaatctactaaacggagagattagatcactgaatcaa 240
|||||
Db 181 taccagatctcaaacggagatgaatctactaaacggagagattagatcactgaatcaa 240
QY 241 tcttcaattccggtactattacacccactagagagagattcaaaagaatccggtccgc 300
|||||
Db 241 tcttcaattccggtactattacacccactagagagagattcaaaagaatccggtccgc 300
QY 301 ttcaagaatattgtcaaatcggtagcgaacacgtgttttactgttttcaaacgggaat 360
|||||
Db 301 ttcaagaatattgtcaaatcggtagcgaacacgtgttttactgttttcaaacgggaat 360
QY 361 ctcgagcaaacacagactcgtgttctcatcacaaagccctcaaaatgtcaaaagaagaca 420
|||||
Db 361 ctcgagcaaacacagactcgtgttctcatcacaaagccctcaaaatgtcaaaagaagaca 420
QY 421 agacggttattctagtagcagcgtgatcatgtttgtttgtttgttaaccagaacacg 480
|||||
Db 421 agacggttattctagtagcagcgtgatcatgtttgtttgtttgttaaccagaacacg 480
QY 481 gattatattcgtttcaaaaacatgattgtgtgttaaccgaacggcgttttctatttc 540
|||||
Db 481 gattatattcgtttcaaaaacatgattgtgtgttaaccgaacggcgttttctatttc 540
QY 541 cgggttcaatgatccgagcgtctgctcaatcagcgtttgtgtgtgtgtgtgtgtac 600
|||||
Db 541 cgggttcaatgatccgagcgtctgctcaatcagcgtttgtgtgtgtgtgtgtgtac 600
QY 601 cgggttcaacggaagaaggatggcattctctaccgtttaaataacggcgttaatttgaga 660
|||||

Db 601 cgggttcaacggaagaaggatggcattctctaccgtttaaataacggcgttaatttgaga 660
QY 661 ctacgaaaaattttgataaaattccgcgcatcaatttatacgcggagatggaattggcg 720
|||||
Db 661 ctacgaaaaattttgataaaattccgcgcatcaatttatacgcggagatggaattggcg 720
QY 721 gtggaattgttttctcctcttctgactgttccattaaaccatcaatcaatctcaagaaaaac 780
|||||
Db 721 gtggaattgttttctcctcttctgactgttccattaaaccatcaatcaatctcaagaaaaac 780
QY 781 gagatgtaggattatccggtgtggaagacgtcgagagataattttatccggtgagaatga 840
|||||
Db 781 gagatgtaggattatccggtgtggaagacgtcgagagataattttatccggtgagaatga 840
QY 841 cgggttttataacgagatgcctatcgaaagtgtgtctgattattcaacggttaaggcag 900
|||||
Db 841 cgggttttataacgagatgcctatcgaaagtgtgtctgattattcaacggttaaggcag 900
QY 901 ctttcggaaacgagtcggttttggatcaactcgtttgcccagcctattcttacagatgaat 960
|||||
Db 901 ctttcggaaacgagtcggttttggatcaactcgtttgcccagcctattcttacagatgaat 960
QY 961 ttggtgttactatcaacctctccaaaatgcccgaatctattcttatttagaagatat 1020
|||||
Db 961 ttggtgttactatcaacctctccaaaatgcccgaatctattcttatttagaagatat 1020
QY 1021 tgaagaacaaatgttatgtgtctatgataataataataataataataataaagattctgc 1080
|||||
Db 1021 tgaagaacaaatgttatgtgtctatgataataataataataataataaagattctgc 1080
QY 1081 gattattttagttatttaattagataagaatttcattcttcttcttcttcttcttcttctga 1140
|||||
Db 1081 gattattttagttatttaattagataagaatttcattcttcttcttcttcttcttcttctga 1140
QY 1141 attacagagacatttaccattttcgaactttgaaacattgaaaaataagaatgaatttaaa 1200
|||||
Db 1141 attacagagacatttaccattttcgaactttgaaacattgaaaaataagaatgaatttaaa 1200
QY 1201 aaaaaaiaaaaaa 1214
|||||
Db 1201 aaaaaaiaaaaaa 1214
RESULT 2
AAF59787
ID AAF59787 standard; cDNA; 1518 BP.
XX
AC AAF59787;
XX
DT 04-MAY-2001 (first entry)
XX
DE Arabidopsis thaliana M8 homeodomain protein cDNA.
XX
KW Homeodomain-like; homeobox gene; M8; plant differentiation;
KW MSH; many shoot; expression control; adventitious bud formation;
KW branching induction; cytokinin independent; transgenic plant; crop;
KW agriculture; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200107618-A1.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-JP04904.
XX
PR 22-JUL-1999; 99JP-0207995.
XX
PA (SUNR) SUNTORY LTD.
XX
PI kakimoto T;
XX
DR WPI; 2001-182796/18.

DR P-PSDB; AAB60643.
 XX Arabidopsis thaliana originated homeobox genes encoding proteins
 PT participating in differentiation with ability to promote adventitious
 PT bud formation and branching induction, applicable in improving plant
 PT cells and plants e.g. for crops -
 XX
 XX Claim 6; Page 25-28; 34pp; Japanese.
 XX
 XX The invention relates to two novel Arabidopsis thaliana proteins M6
 CC and M8 (AAB60642 and AAB60643, respectively), and to the cDNAs
 CC encoding them (AAB60642 and AAB60643). The proteins contain homeodomain-
 CC like sequences and participate in differentiation, having the ability
 CC to promote adventitious bud formation and branching induction by
 CC controlling expression of the many shoot (MSH) gene without the
 CC requirement for cytokinin. The invention also relates to expression
 CC vectors and host cells which contain M6 or M8 nucleic acid sequences,
 CC the recombinant production of the M6 or M8 protein, a plant or plant
 CC cell transfected with the M6 or M8 gene, and methods for inducing
 CC differentiation, adventitious bud formation and/or plant branching via
 CC expression of the transfected M6 or M8 gene. Nucleic acids encoding the
 CC homeodomain proteins may be used to generate transgenic plants with
 CC improved adventitious bud formation and branching for use e.g., as
 CC agricultural crops. The present sequence represents cDNA encoding the
 CC Arabidopsis thaliana M8 protein.
 XX
 XX Sequence 1518 BP; 425 A; 389 C; 279 G; 425 T; 0 other;
 SQ
 Query Match 12.8%; Score 155.2; DB 22; Length 1518;
 Best Local Similarity 63.3%; Pred. No. 7e-21;
 Matches 238; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
 QY 34 tcattgtctctcaacaaattggccagcatgtttcaaatccaaaccccttgcaacaata 93
 DB 150 ttatgtctctctcaacaaattggccagcatgtttcaaatccaaaccccttgcaacaata 209
 QY 94 atcatcatcatcaacatgaatgcagatctcattcttctgcttcttccacgcatcttctcctt 153
 DB 210 aatggcaacatgacatcaatctctctcttcttcttcttcttcttcttcttcttcttctt 269
 QY 154 tcatctctctcttctctcagatcgatgataccagatcttcaaatccaaaccccttgcaacaata 213
 DB 270 tctctcagatgtgaatgcagatgagagtcagatgagcgaacaaacaaacgaatggaatccaaagc 329
 QY 214 cgagcagatgagatcgaatcgaatcttcaatctcgaatcttcaatcttcaatcttcaatctag 273
 DB 330 cagagcagatcggatcgaatcgaatcttcaatcttcaatcttcaatcttcaatcttcaatcag 389
 QY 274 agagattcaaaagaatccggtatccggtcttcaagaatattgtcaaatccggtgagcgaacagc 333
 DB 390 agagattcaggaggattagggtcagcttcaagaatattgtcaaatccggtgagcgaacagc 449
 QY 334 tgtttactggtttcaaaacggaaatctcgagcaaaacaaacagcttctggttcaatcaaca 393
 DB 450 tcttctactggtttcaaaacggaaatctcgagcaaaacaaacagcttctggttcaatcaaca 509
 QY 394 aaagccctcaaaatgctc 409
 DB 510 accactccaaacactc 525
 RESULT 3
 AAS02222
 ID AAS02222 standard; cDNA; 1367 BP.
 XX
 AC AAS02222;
 XX
 XX 16-JUL-2001 (first entry)
 XX
 DE Corn Wuschel (WUS) cDNA from clone p0058_chpab57r.
 XX
 KW Wuschel; WUS; apical meristem; organ formation; homeodomain protein; ss;

transcriptional regulator; proliferation; corn; soybean; wheat; rice;
 alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
 plant transformation; transgenic plant; p0058_chpab57r.
 OS
 XX Zea mays.
 XX
 FH Key Location/Qualifiers
 FT CDS 202..963
 FT /*tag= a
 FT /product= "Corn WUSCHEL protein"
 XX
 XX WO200123575-A2.
 XX
 XX OS-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-US26648.
 XX
 XX 30-SEP-1999; 99US-0157216.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Cahoon RE, Lowe K;
 XX
 XX WPI; 2001-258137/26.
 XX
 XX P-PSDB; AAU01525.
 XX
 XX Novel nucleic acid fragments encoding WUS proteins useful for
 PT transiently modulating WUS protein level in plant cells, as probes for
 PT genetically and physically mapping WUS genes and as markers -
 XX
 XX Claim 4; Page 49-50; 61pp; English.
 CC The sequence represents cDNA from a clone, which encodes a Wuschel (WUS)
 CC protein. The WUS protein plays a key role in initiation and maintenance
 CC of the apical meristem of a plant which contains stem cells, allowing for
 CC organ formation. WUS encodes a homeodomain protein which is thought to
 CC function as a transcriptional regulator. The DNA or its complement is
 CC useful for transiently modulating the level of WUS protein in a plant
 CC cell and meristem proliferation can be induced through its addition. This
 CC involves transforming a plant with WUS DNA to produce a transformed
 CC meristem which is grown into a regenerated plant (corn, soybean, wheat,
 CC rice, alfalfa, sunflower, canola or cotton). The protein and DNA are
 CC useful in the identification of sequences that affect WUS level or enzyme
 CC activity, or that encode homologous proteins from the same or other plant
 CC species. The polypeptides can then be used to prepare antibodies.
 CC Fragments of the DNA may be used as probes to map the genes that they are
 CC part of and as markers for traits linked to those genes, which is useful
 CC in plant breeding to develop lines with desired phenotypes. The nucleic
 CC acid sequences are useful to create transgenic plants in which the WUS
 CC polypeptides are present in higher or lower levels than normal or in cell
 CC types or developmental stages in which they are not normally found.
 XX
 XX Sequence 1367 BP; 291 A; 431 C; 394 G; 251 T; 0 other;
 SQ
 Query Match 5.7%; Score 69.8; DB 22; Length 1367;
 Best Local Similarity 63.3%; Pred. No. 0.00015;
 Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
 QY 196 cgagatggaatcctaaaccggagcagattagtagtactcgaatcaattcttcaatccggtga 255
 DB 254 cgcggtggaaccccaagccgagcagatcactcctcagtcctcctcctcctcctcctcctc 313
 QY 256 ctattaaacccactagagagagattcaagaatccggtatccggttccaaagaatctgctc 315
 DB 314 tggtagaaccccgcaagcagcagcgttcgcgaatcccgagcgtctgctgagcgtctcgcg 373
 QY 316 aaatcgggtgacgcaaacgtgttttactggtttcaaaaacccggaatctctg 364
 DB 374 ccgtggcgagcgcgaacgtcttcttactggttccagaaacccgcgctcccg 422
 RESULT 4

AAC38580
ID AAC38580 standard; DNA; 1326 BP.
XX AC AAC38580;
XX XX
DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21488.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.


```
Qy 296 ccggttcaagaataatgtcaataatcggtgacgcaaacggttttactgtgttccaaacccg 355
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 139 acagctgagcgatgtggaagattgaagcgaagaatgtgttactgtgttcagaacca 198
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Qy 356 gaaatctcgagcaaaaacacaaagc 378
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 199 caaagccagagaagaagcgagaagc 221

RESULT 6
AAS02225
ID AAS02225 standard; cDNA; 506 BP.
AC AAS02225;
XX
XX
DT 16-JUL-2001 (first entry)
DE Soybean Wuschel cDNA from contig of scrlc_pk001_d2 and ses4d_pk0033_c8.
XX
XX Wuschel; WUS; apical meristem; organ formation; homeodomain protein; ss;
KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
KW plant transformation; transgenic plant; scrlc_pk001_d2; ses4d_pk0033_c8.
XX
XX Glycine max.
XX
XX Key Location/Qualifiers
XX CDS 21..425
XX FT /*tag= a
XX FT /product= "Corn WUSCHEL protein"
XX
XX WO200123575-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US266648.
XX
XX 30-SEP-1999; 99US-0157216.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Cahoon RE, Lowe K;
XX
XX WPI; 2001-258137/26.
XX
XX P-PSDB; AAU01528.
XX
XX Novel nucleic acid fragments encoding WUS proteins useful for
XX transiently modulating WUS protein level in plant cells, as probes for
XX genetically and physically mapping WUS genes and as markers
XX
XX Claim 4; Page 54; 6lpp; English.
XX
XX The sequence represents cDNA from a contig of clones, which encodes a
XX Wuschel (WUS) protein. The WUS protein plays a key role in initiation and
XX maintenance of the apical meristem of a plant which contains stem cells,
XX allowing for organ formation. WUS encodes a homeodomain protein which is
XX thought to function as a transcriptional regulator. The DNA or its
XX complement is useful for transiently modulating the level of WUS protein
XX in a plant cell and meristem proliferation can be induced through its
XX addition. This involves transforming a plant with WUS DNA to produce a
XX transformed meristem which is grown into a regenerated plant (corn,
XX soybean, wheat, rice, alfalfa, sunflower, canola or cotton). The protein
XX and DNA are useful in the identification of sequences that affect WUS
XX level or enzyme activity, or that encode homologous proteins from the
XX same or other plant species. The polypeptides can then be used to prepare
XX antibodies. Fragments of the DNA may be used as probes to map the genes
XX that they are part of and as markers for traits linked to those genes,
XX which is useful in plant breeding to develop lines with desired
XX phenotypes. The nucleic acid sequences are useful to create transgenic
XX plants in which the WUS polypeptides are present in higher or lower
XX levels than normal or in cell types or developmental stages in which they
XX are not normally found.
XX
```

```
SQ Sequence 506 BP; 159 A; 102 C; 120 G; 122 T; 3 other;
    Query Match 5.0%; Score 60.8; DB 22; Length 506;
    Best Local Similarity 56.5%; Pred. No. 0.0073;
    Matches 113; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 199 gatggaatcctaaaccgagcagattaggtactcgcgaatcaatcttcaattccggtacta 258
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 97 gatggagtctctacaaagagcaaatagacatgttggagaacttttacaagcagggaataa 156
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Qy 259 ttaaccacacctagagagagattcaagaagaatccggatccggtctcaagaatatgtgtcaaa 318
    || | || | || | || | || | || | || | || | || | || | || | || | || |
Db 157 ggactccagcactgagcaaatatacaacagattactctaggcttagggcttatgtttaca 216
    || | || | || | || | || | || | || | || | || | || | || | || | || |
Qy 319 tcggtgacgcaaacgtgttttactgtgttcaaaaaccggaatctcgagcaaaacacaaagc 378
    || | || | || | || | || | || | || | || | || | || | || | || | || |
Db 217 tcgagggaaaaaatgtcttactgtgttcaaaatcacaagcgcccaagacagaagc 276
    || | || | || | || | || | || | || | || | || | || | || | || | || |

Qy 379 tcgtgtttcatcaaaaagc 398
    | | | | |
Db 277 tcaagcagaagaacaaagc 296

RESULT 7
AAS02226
ID AAS02226 standard; cDNA; 844 BP.
XX
XX AC AAS02226;
XX
XX DT 16-JUL-2001 (first entry)
XX
XX DE Soybean Wuschel (WUS) cDNA from clone scrlc_pk001_d2.
XX
XX KW Wuschel; WUS; apical meristem; organ formation; homeodomain protein; ss;
XX transcriptional regulator; proliferation; corn; soybean; wheat; rice;
XX alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
XX KW plant transformation; transgenic plant; scrlc_pk001_d2.
XX
XX OS Glycine max.
XX
XX FH Key Location/Qualifiers
XX CDS 8..661
XX FT /*tag= a
XX FT /product= "Corn WUSCHEL protein"
XX FT /partial
XX FT /note= "No start codon"
XX
XX WO200123575-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US266648.
XX
XX 30-SEP-1999; 99US-0157216.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Cahoon RE, Lowe K;
XX
XX WPI; 2001-258137/26.
XX
XX P-PSDB; AAU01529.
XX
XX Novel nucleic acid fragments encoding WUS proteins useful for
XX transiently modulating WUS protein level in plant cells, as probes for
XX genetically and physically mapping WUS genes and as markers
XX
XX Claim 4; Page 55; 6lpp; English.
XX
XX The sequence represents cDNA from a clone, which encodes a Wuschel (WUS)
XX protein. The WUS protein plays a key role in initiation and maintenance
XX of the apical meristem of a plant which contains stem cells, allowing for
XX organ formation. WUS encodes a homeodomain protein which is thought to
```


xx 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125786.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140822.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

```
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 03-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.9%; Score 59.4; DB 21; Length 945;
Best Local Similarity 56.3%; Pred. No. 0.014;
Matches 111; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 190 ctaaacgagatgaatccctaaacggagcagattagtagctacgaaatcctcaatt 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 ctgacagaggtggagctccctacgcagctccagcttcgagattctggaacgtatctcgc 484

QY 250 ccggtactattaaacccacctagagagagattcctaaagaatccggctccagaat 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 aaggcaggggaacacccgagcaagcagagatcaagaacataacagagagctgagccaac 544

QY 310 atggtcaaatcgtgacgcgaacggtgtttactggtttcctaaacccggaatctcagcaa 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 acggccagattgctgaacaaaatgtctataactgttccagaaacggagcgtgctgatcca 604

QY 370 aacacaagcttcgtgtt 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 agaggaagcagcatggt 621

RESULT 12
AAS02227
ID AAS02227 standard; cDNA; 741 BP.
XX
XX AAS02227;
XX
XX 16-JUL-2001 (first entry)
XX Soybean Wuschel (WUS) cDNA from clone ses4d_pk0033_c8.
DE Wuschel; WUS; apical meristem; organ formation; homeodomain protein; ss;
KW transcriptional regulator; proliferation; corn; soybean; wheat; rice;
KW alfalfa; sunflower; canola; cotton; enzyme activity; genetic marker;
KW plant transformation; transgenic plant; ses4d_pk0033_c8.
XX
OS Glycine max.
```

```
XX Key Location/Qualifiers
FH CDS 27..665
FT /*tag= a
FT /product= "Corn WUSCHEL protein"
XX
XX WO200123575-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26648.
XX
XX 30-SEP-1999; 99US-0157216.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Cahoon RE, Lowe K;
XX
XX WPI; 2001-258137/26.
XX P-PSDB; AAU01530.
XX
XX Novel nucleic acid fragments encoding WUS proteins useful for
XX transiently modulating WUS protein level in plant cells, as probes for
XX genetically and physically mapping WUS genes and as markers -
XX
XX Claim 4; Page 56; 61pp; English.
XX
XX The sequence represents cDNA from a clone, which encodes a Wuschel (WUS)
XX protein. The WUS protein plays a key role in initiation and maintenance
XX of the apical meristem of a plant which contains stem cells, allowing for
XX organ formation. WUS encodes a homeodomain protein which is thought to
XX function as a transcriptional regulator. The DNA or its complement is
XX useful for transiently modulating the level of WUS protein in a plant
XX cell and meristem proliferation can be induced through its addition. This
XX involves transforming a plant with WUS DNA to produce a transformed
XX meristem which is grown into a regenerated plant (corn, soybean, wheat,
XX rice, alfalfa, sunflower, canola or cotton). The protein and DNA are
XX useful in the identification of sequences that affect WUS level or enzyme
XX activity, or that encode homologous proteins from the same or other plant
XX species. The polypeptides can then be used to prepare antibodies.
XX Fragments of the DNA may be used as probes to map the genes that they are
XX part of and as markers for traits linked to those genes, which is useful
XX in plant breeding to develop lines with desired phenotypes. The nucleic
XX acid sequences are useful to create transgenic plants in which the WUS
XX polypeptides are present in higher or lower levels than normal or in cell
XX types or developmental stages in which they are not normally found.
XX
XX Sequence 741 BP; 238 A; 155 C; 156 G; 192 T; 0 Other;
```

```
Query Match 4.8%; Score 57.8; DB 22; Length 741;
Best Local Similarity 57.5%; Pred. No. 0.028;
Matches 104; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 199 gatggaatcttaaacggagcagattagtagctacgaaatcctcaatcctcgggtacta 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 ggtggagtcctacaaggagcaaatagacatgttagagaacctttacaagcagggaataa 147

QY 259 ttaacccacctagagagagattcaagaatccggatccgcttcaagaatagtgtcaaa 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 ggactccacgactgagcaataacacagattaccttagctaggctcagggcttaggtcaca 207

QY 319 tcggtgacgcaaacgtgttttactggtttcaaaacccggaaatctcgcgaaacacagc 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 tcgagggaagaatgtcttctactggtttcaaaatcacaaagctcgtcaagacagagc 267

QY 379 t 379
Db 268 t 268

RESULT 13
AAS02227
```


CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO
CC ccp.wipo.int/pub/published_pct_sequences
XX
SO Sequence 6509 BP; 1935 A; 113 C; 1338 G; 3123 T; 0 other;

Query Match	4.6%	Score 55.8;	DB 24;	Length 6509;
Best Local Similarity	53.4%;	Pred. No. 0.083;		
Matches 117;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps 0;
Qy	996	atctattatcttattgagaagatattgaaagcgaatgtattggctgcatggataaataat	1055	
Db	3673	ATAAACACCTAAATATTACATATTAAAAAAAATAATTTAAAAAAAACATATAT	3614	
Qy	1056	taataataataaagaattctcgatttatttagttattaatagataagaatttcat	1115	
Db	3613	AATACAAAAATAAAAAATAAATAATCTAAATAATTTAAATATCTATATAAAATCTTTAAC	3554	
Qy	1116	tcttattcttttaatttatgaaacatttaccagacatttaccatttctcgagccttga	1175	
Db	3553	CTCAATACTATTATATATATCATATAAATCTTTTCATTTTATTTTCTAAAAAATATTT	3494	
Qy	1176	aataagaatgaaatttaagtttaaaaaaataaaaaaa	1214	
Db	3493	AACACCTCTTAATTTTATATATAAAAAATAATATAAAATAA	3455	

RESULT 15
ABL34155/C
ID ABL34155 standard; DNA: 15548 BP.
XX
AC ABL34155;

26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 2128.

Human: immune system disease; cytosine methylation; antiasthmatic;
antiartherosclerotic; antiamebic; cytostatic; nootropic;
neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
antirheumatic; antarthritic; antidiabetic; antipsoriatic;
antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
gene: ds.

XX	Homo sapiens.
OS	
XX	
XX	
PN	WO200200928-A2.
XX	
XX	
XX	03-JAN-2002.
PD	
XX	
XX	02-JUL-2001; 2001WO-EP07537.
PF	
XX	
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	
XX	01-SEP-2000; 2000DE-1043826.
XX	
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
XX	
DR	WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX
PS Claim 1; SEQ ID NO 2128; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/vulcerative bowel
CC diseases. The present sequence is a gene of the invention

	Query Match	4.68;	Score 55.4;	DB 24;	Length 15548;
	Best Local Similarity	50.68;	Pred. No. 0.11;	131;	Indels 0;
	Matches 134;	Conservative 0;	Mismatches		Gaps
Qy	950	tacagatgaatttggtgttacttatcaacctctccaaatgcgccaattctattctcttat	1009		
Db	12024	TAAAAATAAAAATATATATAACATATAAAAAATAAAATATATATAACATATAAAAAATAAAAT	11965		
Qy	1010	ttgaagatatattgaaagcaaatgttatggctgctatggataaataattataataacaataa	1069		
Db	11964	ATATATAACATATATAAAATAACATATAAAAATAAAATATAAAATAAAATATAAAANATAA	11905		
Qy	1070	aagatttctcgagatttatttgattattaattagatagaagaatttcattctctcttttaa	1129		
Db	11904	AATAAATATAAAATATAAAAAATAAAATAAAATATAAAATATAAAATAAAATAAAATAAA	11845		
Qy	1130	atttatgaacaattttacaggacattttacatttttcgagactttcgaaataaagaatgaa	1189		
Db	11844	ATATATAAAATATAAATATAAAATAAAATATAAAATAAAATATAAAATAAAATATAAA	11785		
Qy	1190	ttaggttataaaaaaataaaaaa	1214		
Db	11784	ATAAAATATAAAATATAAAATATAA	11760		

Search completed: August 27, 2002, 21:33:26
Job time: 3853 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2002, 20:23:13 ; Search time 1598.05 Seconds
(without alignments)
10253.311 Million cell updates/sec

Title: US-09-787-737-1

Perfect score: 1214

Sequence: 1 ctttagctctcgattatcat.....ttaaaaaaaaaaaaaaaaaa 1214

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpi:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pln:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	322.2	26.5	327	9	AV556647 AV556647
C 2	305.6	25.2	336	9	AV557790 AV557790
C 3	159	13.1	207	12	BH253315 SALX_0147
4	156.4	12.9	446	12	BH634225 SALX_0449
5	155.2	12.8	530	9	AV549782 AV549782
6	132.6	10.9	604	12	BH571919 BCGY216TR
7	109.8	9.0	523	12	AQ844430 an37c06 J
8	101	8.3	594	9	AW981538 PC13G09 P
9	93.8	7.7	852	12	BH607214 BOHGN75TF
10	80	6.6	610	10	BG442689 GA_Ea001
11	78.4	6.5	506	10	BM086304 sah38c06
C 12	77	6.3	503	9	AV537767 AV537767
13	72.2	5.9	723	12	CNS021AR
14	71.4	5.9	510	10	BE472479 EST417332
15	71.4	5.9	928	12	CNS00DKY
16	70.4	5.8	576	12	AQ272170 nbxb0027J
17	69.4	5.7	831	12	AQ364605 nbxb0061E

18	67.8	5.6	556	12	AQ795517
19	66.6	5.5	477	10	BI209277
20	66.6	5.5	557	10	BG134747
21	66.6	5.5	627	10	BI204369
22	66	5.4	843	12	CNS00CS1
C 23	65.2	5.4	390	12	B23895
C 24	65.2	5.4	541	12	CNS00P0R
C 25	65.2	5.4	1101	12	CNS00B0I
C 26	64.4	5.3	1200	12	CNS016C0
27	64.2	5.3	1043	12	CNS0145P
C 28	63.6	5.2	928	12	CNS00DKY
29	63	5.2	589	12	CNS0301Z
C 30	63	5.2	1101	12	CNS017V2
C 31	62.8	5.2	637	12	CNS036CC
C 32	62.8	5.2	1029	12	CNS012GM
C 33	62.8	5.2	1101	12	CNS016LI
34	62.6	5.2	551	9	AW624754
35	62.6	5.2	560	10	BG134403
36	62.6	5.2	751	10	BI930760
37	62.2	5.1	472	10	BG132124
38	62	5.1	462	9	AW624610
39	62	5.1	565	9	AW945094
40	62	5.1	924	12	CNS07A5L
C 41	61.8	5.1	1101	12	CNS0039G
C 42	61.6	5.1	702	12	B28518
C 43	61.6	5.1	935	12	CNS00JKN
C 44	61.6	5.1	1101	12	CNS0021J
C 45	61.4	5.1	1203	12	CNS015WU

ALIGNMENTS

RESULT 1

AV556647/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Score

26.5%

Score

322.2

DB 9

Length

327

AV556647

Arabidopsis thaliana

CDNA clone SQ048H01F 3'

327 bp mRNA linear

EST 07-SEP-2000

Arabidopsis thaliana

CDNA clone SQ048H01F 3'

327 bp mRNA linear

EST 07-SEP-2000

Arabidopsis thaliana

CDNA clone SQ048H01F 3'

327 bp mRNA linear

EST 07-SEP-2000

Arabidopsis thaliana

CDNA clone SQ048H01F 3'

327 bp mRNA linear

EST 07-SEP-2000

Best Local Similarity 99.1%; Pred. No. 3.2e-38; Matches 324; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 805 agacgtcagagataatgtttatccggtgagaaacagcggtgtttatcaacgagatcccta 864
 |||||
 Db 327 AGACGTCGAGATATGTTTATCCGGTGAGAAATGACGGTGTATTATCAAGAGATGCCCTA 268
 |||||
 QY 865 tcgaagtagtctgtgattattcaacgcttaagcgagctttcggaaacgagatgcgcttttga 924
 |||||
 Db 267 TCGAAGTAGTCTGGATTATTCAACGTTTAAGCAGCTTTCGGAACGATGCCGTTTGA 208
 |||||

QY 925 tcaactcgtttggcgcgtattcttcaacgagatgaatttggttacttataacacctctcc 984
 |||||
 Db 207 TCAACTCGTGTGGCCAGCCTATTCTTTACAGATGAATTTGGTGTACTTTATCAACCTCTCC 148
 |||||
 QY 985 aaaaagcgaactctattctatttagaagatatgaaaacgaatgttatggtgcta 1044
 |||||
 Db 147 AAATGGCGCAATCTATTATCTATTATTTAGAAGATATTGAAAGCAAAATGTTATGGTGCTA 88
 |||||

QY 1045 tggataaatttaataataataaagattctcgcgattatttagttatttaattagat 1104
 |||||
 Db 87 TCGATAAATTAATAATAATAAAGATTTTCGGCGGTTTATTAGTTATTAAATTATAT 28
 |||||

QY 1105 aagaatttcattcttatctatttaaat 1131
 |||||
 Db 27 AAGATTTTCATTTCTTATCTTTAAAT 1

RESULT 2
 AV557790/c
 LOCUS AV557790 Arabidopsis thaliana green silicles Columbia Arabidopsis
 DEFINITION thaliana cDNA clone SQ080b05F 3', mRNA sequence.
 ACCESSION AV557790
 VERSION AV557790.1 GI:8729205
 KEYWORDS EST.
 SOURCE chae cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 336)
 Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
 source Location/Qualifiers
 1..336
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ080b05F"
 /clone_lib="Arabidopsis thaliana green silicles Columbia"
 /tissue_type="green silicles"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 125 a 59 c 48 g 104 t
 ORIGIN

Query Match 25.2%; Score 305.6; DB 9; Length 336;
 Best Local Similarity 98.2%; Pred. No. 8.6e-36;
 Matches 330; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 818 taatgtttatccggtgagatgacggtgtttatttaacgagatccctatcgaagtatgctc 877
 |||||

Db 336 TAAATGTTTATCCGGTGAGAAATGACGGTGTATTATTAACGAGATGCCTATCGAAGTAGTGTC 277
 |||||
 QY 878 tggattattcaacgcttaagcgagctttcggaaacgagatgcgcttttgcatacactcgtttgg 937
 |||||
 Db 276 TGGATTATTCAACGTTTAAGCAGCTTTTCGGAAGCGATGCGGTTTGTATCAACTCGTTTGG 217
 |||||
 QY 938 ccagcctattctacagatgaatttggtgttacttatacaacctctccaaaatggcgcaat 997
 |||||
 Db 216 CCAGCGCTATTCTTACAGATGAATTTGGTGTACTTATCAACCTCTCCAAAATGGCGCAAT 157
 |||||
 QY 998 ctattatctatttagaagatatgaaaacgaatgttatggtgctatgsgataaata 1057
 |||||
 Db 156 CTATTATCTTATTAGAAGATATTGAAAGCAAAATGTTATGGTGCTATGGATAAATATTA 97
 |||||
 QY 1058 atataataataaagattctcgcgattatttagttatttaatt-agaataagaattccatt 1116
 |||||
 Db 96 ATATTATAATAAAGATTTCTGCGATTTATTAGTTATTACTTCATATAGAATTTTCATT 37
 |||||
 QY 1117 tcttctcttttaaat-ttatgaacaattttacaggac 1151
 |||||
 Db 36 TCTTATCTTTTAAATATTATGAACAATTTACAGGAC 1

RESULT 3
 BH253315/c
 LOCUS BH253315 207 bp DNA linear GSS 28-NOV-2001
 DEFINITION SALK_014799 Arabidopsis thaliana TDNA insertion lines Arabidopsis
 thaliana genomic clone SALK_014799, DNA sequence.

ACCESSION BH253315
 VERSION BH253315.1 GI:17140815
 KEYWORDS GSS.
 SOURCE chae cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 207)
 Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
 C., Heller,C., Kim,C.J., Jeske,A., Koesema,E., Meyers,M.C., Parker
 H., Prednis,L., Shinn,P., Stevenson,D.K., Zimmerman,J. and Ecker
 J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within an annotated exon of At5g45980.
 Class: TDNA tagged.

FEATURES
 source Location/Qualifiers
 1..207
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_014799"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 55 a 51 c 46 g 55 t
 ORIGIN

Query Match 13.1%; Score 159; DB 12; Length 207;
 Best Local Similarity 100.0%; Pred. No. 3.6e-14;


```

Db 392 TCTTCTACTGGTTCCAAAACCGTAAGTCCCGTAGTAGTAACACAAACTCCGCGCTCCTCCACA 451
Qy 394 aaagccctaaatgtc 409
| | | | |
Db 452 ACCACTCCAAACACTC 467

RESULT 6
LOCUS BH571919 604 bp DNA linear GSS 14-DEC-2001
DEFINITION BOGYZ16TR BOGY Brassica oleracea genomic clone BOGYZ16, DNA
sequence.
ACCESSION BH571919
VERSION BH571919.1 GI:17823758
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 604)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGYZ16TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
Location/Qualifiers
source 1..604
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGYZ16"
/clone_lib="BOGY"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 168 a 153 c 117 g 166 t
ORIGIN

Query Match 10.9%; Score 132.6; DB 12; Length 604;
Best Local Similarity 74.0%; Pred. No. 1.7e-10;
Matches 168; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 174 gatcgataccatctcctaaccgagatgaatcctaaccgagagagagattagatactc 233
| | | | |
Db 152 GAGAGAGTCCAGAGCCAAACCAAGATGGTAATCCGGAAGCAGAGAGATTCGTGTACTT 211
| | | | |

Qy 234 gaatcaatcttcattccggtactattaaacccctagagagagattcaagaatccgg 293
| | | | |
Db 212 GAACCAATCTTCAACTCAGGATGGTGAATCCACCAAGAGAGAGATTCGAAGGATTAGG 271
| | | | |

Qy 294 atccggcttcaagaatatgttcaaatcggtagcgaacacgtgtttactggtttcaaac 353
| | | | |
Db 272 GCTCAGCTTCAAGAATAACGCCAAGTTGGTGAGGCCAACGCTTCTTATTGTTCCAAAC 331
| | | | |

Qy 354 cgaataatctcagcaaacacaaagcttcgtgttcacacaaagccc 400
| | | | |
Db 332 CGAAGTCTCGTAGCAAAACAAATAACGCTCTCTCTATTAACATC 378
| | | | |

RESULT 7
LOCUS AQ844430 523 bp DNA linear GSS 03-NOV-1999
DEFINITION an37c06 JM101 filtered library Zea mays genomic, DNA sequence.
ACCESSION AQ844430

```

```

VERSION AQ844430.1 GI:6202918
KEYWORDS GSS.
SOURCE ze mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 523)
AUTHORS Rabinowicz,P.D., Schutz,K., Dedhia,N., Yordan,C., Parnell,L.D.,
Stein,L., McCombie,W.R. and Martienssen,R.A.
TITLE Differential methylation of genes and retrotransposons allows
shotgun sequencing of the maize genome
JOURNAL Nat. Genet. 23, 305-308 (1999)
COMMENT Contact: Martienssen RA
Cold Spring Harbor Laboratory
1 Bungtown Rd., Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8322
Fax: 516 367 8369
Email: martiens@eshl.org
Seq primer: forward
Class: shotgun.

FEATURES
Location/Qualifiers
source 1..523
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="JM101 filtered library"
/note="Organ: immature ears; Vector: M13; Site_1: Xba I;
DNA prepared from purified nuclei was digested with the
methylation insensitive enzyme Spe I, size fractionated to
enrich for the 0.5 to 4 kbp fraction, ligated into Xba I
digested M13 vector and electroporated into E.coli JM101."
BASE COUNT 92 a 163 c 170 g 98 t
ORIGIN

Query Match 9.0%; Score 109.8; DB 12; Length 523;
Best Local Similarity 70.3%; Pred. No. 4e-07;
Matches 147; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 174 gatcgataccatctcctaaccgagatgaatcctaaccgagagagattagatactc 233
| | | | |
Db 185 GAGCGCCCGCCGATCGGAAGCCGGTGGAAACCCCGAGCGGAGATCCCGGATTCTG 244
| | | | |

Qy 234 gaatcaatcttcattccggtactattaaacccctagagagagattcaagaatccgg 293
| | | | |
Db 245 GAGGCATCTTCAACTCCGCGATGGTCAACCCGCGCGCAGAGATCCCGCATCCG 304
| | | | |

Qy 294 atccggcttcaagaatatgttcaaatcggtagcgaacacgtgtttactggtttcaaac 353
| | | | |
Db 305 ATGCGCTGCAGCAGTACGGCGCAGGTGCGGAGCCCAACGCTTCTTACTGTTCCAGAAC 364
| | | | |

Qy 354 cgaataatctcagcaaacacaaagcttcg 382
| | | | |
Db 365 CGAAGTCCCGCTCCAAGAACAAAGCTCCG 393
| | | | |

RESULT 8
LOCUS AW981538 594 bp mRNA linear EST 02-JUN-2000
DEFINITION FC13G09 Pine Triplex pollen cone library Pinus taeda cDNA clone
FC13G09, mRNA sequence.
ACCESSION AW981538
VERSION AW981538.1 GI:8173101
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
REFERENCE 1 (bases 1 to 594)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project

```

JOURNAL
COMMENT

Unpublished (1999)
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.
Location/Qualifiers

FEATURES

source

1. .594
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="PC13G09"
/clone_lib="Pine Triplex pollen cone library"
/dev_stage="Immature"
/lab_host="E. coli BM25.8"
/note="Organ: pollen cone; Vector: Lambda Triplex; Site_1:
SfiI (A); Site_2: SfiI (B); Immature pollen cones were
collected in the early spring, frozen and used for mRNA
isolation. The SMART-PCR method (Clontech) was used to
prepare a library from 1 ug total RNA, using the lambda
Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."

BASE COUNT

186 a 121 c 121 g 166 t

ORIGIN

Query Match 8.3%; Score 101; DB 9; Length 594;
Best Local Similarity 68.3%; Pred. No. 7.4e-06;
Matches 140; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 178 ggataccagatcctaaacggagatgaatcctaaccggagcagataggatcactcgaat 237
Db 206 GCACACAGATCCAAAGCCAGATGGAATCCGACGCCCTGAACAGGTTTCAGATTTTGGAA 265
QY 238 caatctcaattccggtactatataccacactagagagagattcaagaatccggatcc 297
Db 266 AGATATTCAATTCGTGTAAGGTGAACCTCCACGACAGGAATTAATAAATCACCACCTC 325
QY 298 ggcctcaagaatattggtcaaatcggtacgcaaacggttttactggtttcaaaaccgga 357
Db 326 AGCTGCAAGAATTCCGAGACGCTTAGGAGGCCAATATCATCTACTGTTCCAGAACCGAA 385
QY 358 aatctcgagcaaacacaaagcttcg 382
Db 386 ACTCCAGATCCAAACAAAAGCAACG 410

RESULT 9

BH607214

LOCUS BH607214 BOHG Brassica oleracea genomic clone BOHGN75, DNA
DEFINITION BOHG75TF BOHG Brassica oleracea genomic clone BOHGN75, DNA
sequence.

ACCESSION

BH607214

VERSION BH607214.1 GI:17859660

KEYWORDS

GSS.

SOURCE

ORGANISM

Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 852)
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOHGN75TF
Contact: Chris Town
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1. .852

/organism="Brassica oleracea"

/strain="TOL000DH3"

/db_xref="taxon:3712"

/clone="BOHGN75"

/clone_lib="BOHG"

/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"

242 a 203 c 175 g 232 t

BASE COUNT

ORIGIN

Query Match 7.7%; Score 93.8; DB 12; Length 852;
Best Local Similarity 67.9%; Pred. No. 7.1e-05;
Matches 131; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 199 gatgaatcctaaacggagcagatgagatgatactagaatcaatctcgaatccggtacta 258
Db 81 GTTGTCTACCTAAACCGGAGCAAAATCTTGATCTCGAATCCATCTTCAACAGTGTATTG 140
QY 259 ttaacccacctagagagagattcaagaatccggtccggtccgaagaatattggtcaaa 318
Db 141 TTAACCCACCAAGATGAACGGTAAGGATAGAAAGATGCTAGAAAATTTGGTGCAG 200
QY 319 tcggtgacgcaaacgtgttttactggtttcaaacccggaaatctcgagcaaacacaagc 378
Db 201 TGGGAGACGCAACGCTCTCTACTGTTTCAAAACCGCGGTCCAGCATCTCGCCGGAGAC 260
QY 379 ttcgtgttcataca 391
Db 261 AGCGGCAACTTCA 273

RESULT 10

BG442689

LOCUS

DEFINITION GA_Ea0017P06f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0017P06f, mRNA sequence.

ACCESSION

BG442689

VERSION

BG442689.1 GI:13352341

KEYWORDS

EST.

SOURCE

ORGANISM

Gossypium arboreum.
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 610)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

Seq primer: TAATACGACTCATTATAGG

High quality sequence stop: 461.

Location/Qualifiers

1. .610

/organism="Gossypium arboreum"

/strain="ARA"

/cultivar="8400"

/db_xref="taxon:29729"

.....

Thu Aug 29 11:18:04 2002

us-09-787-737-1.rst

Page 9

THIS PAGE BLANK (UEPTD)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2002, 20:28:28 ; Search time 49.57 Seconds
(without alignments)
6015.717 Million cell updates/sec

Title: US-09-787-737-1
Perfect score: 1214
Sequence: 1 ctttagctcgtattcat.....ttaaaaaaaaaaaaaaaaaa 1214

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49.4	4.1	3138	1 US-07-867-106-4	Sequence 4, Appl
2	49.4	4.1	6152	4 US-08-973-462-1	Sequence 1, Appl
3	47.8	3.9	1959	3 US-09-166-460-1	Sequence 1, Appl
4	47.8	3.9	1959	4 US-09-361-718-1	Sequence 1, Appl
5	47.6	3.9	1364	1 US-08-265-087-3	Sequence 3, Appl
6	47.6	3.9	1364	1 US-08-621-493-3	Sequence 3, Appl
7	47.6	3.9	1364	2 US-08-965-688-3	Sequence 3, Appl
8	47.6	3.9	1364	4 US-09-260-173-3	Sequence 3, Appl
9	47.2	3.9	19124	2 US-08-487-8268-13	Sequence 13, Appl
10	45.8	3.8	3300	3 US-08-913-842-4	Sequence 4, Appl
C 11	45.4	3.7	19124	2 US-08-487-8268-13	Sequence 13, Appl
12	45.2	3.7	1939	1 US-07-715-7518-2	Sequence 2, Appl
13	45	3.7	2409	4 US-09-293-324C-8	Sequence 8, Appl
C 14	44.8	3.7	6124	4 US-08-213-4198-3	Sequence 3, Appl
15	44.4	3.7	837	4 US-08-998-416-288	Sequence 288, App
16	44.2	3.6	2317	3 US-08-749-522-5	Sequence 5, Appl
17	44.2	3.6	3138	1 US-07-867-106-4	Sequence 4, Appl
18	44.2	3.6	3974	4 US-08-467-504-3	Sequence 3, Appl
19	44	3.6	2406	2 US-08-687-702-2	Sequence 2, Appl
20	43.8	3.6	1117	4 US-09-247-373B-33	Sequence 33, Appl
C 21	43.8	3.6	5852	1 US-07-867-106-2	Sequence 2, Appl
C 22	43.8	3.6	6152	4 US-08-973-462-1	Sequence 1, Appl
23	43.6	3.6	8920	2 US-08-446-855A-1	Sequence 1, Appl
24	43.6	3.6	8920	4 US-09-150-741-1	Sequence 1, Appl
25	43.4	3.6	2058	2 US-08-749-391-1	Sequence 1, Appl
26	43.4	3.6	2058	3 US-09-390-200-1	Sequence 1, Appl
27	43.4	3.6	5852	1 US-07-867-106-2	Sequence 2, Appl

28	43	3.5	1669	4 US-09-461-697-184	Sequence 184, App
29	42.4	3.5	2110	4 US-09-419-459-1	Sequence 1, Appl
30	42.4	3.5	2837	2 US-08-993-228-11	Sequence 11, Appl
31	42	3.5	2836	3 US-08-747-221B-24	Sequence 24, Appl
C 32	42	3.5	2836	3 US-08-747-221B-26	Sequence 26, Appl
C 33	42	3.5	2836	4 US-09-005-051-24	Sequence 24, Appl
C 34	42	3.5	2836	4 US-09-005-051-26	Sequence 26, Appl
35	41.8	3.4	636	4 US-08-998-416-1137	Sequence 1137, Ap
36	41.8	3.4	1582	3 US-08-545-196B-10	Sequence 10, Appl
C 37	41.8	3.4	1582	3 US-08-545-196B-12	Sequence 12, Appl
C 38	41.8	3.4	2448	1 US-08-526-964-2	Sequence 2, Appl
C 39	41.8	3.4	2448	2 US-08-946-617-2	Sequence 2, Appl
C 40	41.8	3.4	2448	3 US-09-031-897-2	Sequence 2, Appl
C 41	41.8	3.4	2663	1 US-08-136-743B-3	Sequence 3, Appl
42	41.8	3.4	4539	1 US-08-119-512-1	Sequence 1, Appl
43	41.8	3.4	4539	1 US-08-488-015B-1	Sequence 1, Appl
44	41.8	3.4	4542	3 US-08-814-412-11	Sequence 11, Appl
C 45	41.8	3.4	5203	4 US-09-257-770-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-07-867-106-4/c
; Sequence 4, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3138 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
US-07-867-106-4

Query Match 4.1%; Score 49.4; DB 1; Length 3138;
Best Local Similarity 57.4%; Pred. No. 0.011;

Qy	1214	a	1214
		-	
Db	1969	A	1969

RESULT 11

```

US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellemans, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

```

	Query Match	3.7%	Score 45.4	DB 2	Length 19124
	Best Local Similarity	55.4%	Pred. No. 0.15		
	Matches 148	Conservative	0	Mismatches 116	Indels 3
	Gaps				
Qy	951	acagatgaatttggtgcttaacttacaacctctcca-aaatgcgcgaatccttattctcttat	1009		
Db	18253	AAATAAGAAATTAAGTATATATTTAAAAATAACAAAGAAATATTTTGTATATATATATAA	18194		
Qy	1010	ttagaagatatgaagcgaatgttatgtgctatgataaataatata-acataataata	1068		
Db	18193	TTTATATTTTAATAGGAACATATATAATATTTGTATTTATAAATAATTTTATATATATATATATATA	18134		
Qy	1069	aaagattctgcgattbattbattgattataatagatagaagaatttcattcttcttatta	1128		
Db	18133	TATTTTTTTTTTCGCTCGGATTTATTCAGAAATTAGAGTAAACAAAAATAAACACATATATATA	18074		
Qy	1129	aatttat-gaacaatttacaggacatttcacattcttcagaccttgcgaataataaagaatga	1187		

Db	18073	AACACATATATAATAACACATTTAAATAATATATATATAAATAATATACACAAATAATATAAATA	18014
Qy	1188	aattaagtgtaaaaaaataaaaaaataaaaaa	1214
Db	18013	AAATTAATCTGCAAAAAAATAATGAAAAAA	17987

RESULT 12

```

US-07-715-751B-2
: Sequence 2, Application US/07715751B
: Patent No. 5391725
: GENERAL INFORMATION:
: APPLICANT: CORUZZI, GLORIA M
: APPLICANT: EDWARDS, JANICE W
: APPLICANT: WALKER, ELSEBETH L
: APPLICANT: BREAKS, TIMOTHY B
: TITLE OF INVENTION: NOVEL ORGAN-SPECIFIC PLANT PROMOTER
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 AVENUE OF THE AMERICAS
: CITY: NEW YORK
: STATE: N.Y.
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/715,751B
: FILING DATE: 19910613
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MISROCK, S. LESLIE
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 3288-017-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1919 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-07-715-751B-2

```

	Query Match	3.7%	Score 45.2	DB 1	Length 1939
	Best Local Similarity 50.5%	Pred. No. 0.09			
	Matches 110	Conservative 0	Mismatches 108	Indels 0	Gaps 0
Qy	997	tctattcttattagaagatttgaaagcaaatgttatggctgatgataaaatt	1056		
Db	1193	TTTATTAGTCATCTATTAATAGAGAAATTCATAATACAAAGTTTGAATTGTGAATTTT	1242		
Qy	1057	aataataataaagatttctgcgatttatttggttattattagataagaatttcatt	1116		
Db	1243	GATTATTAGATACTAACAGAGTGAATTTTAATTATTAGATATTTAATAAAAATATTATT	1302		
Qy	1117	tcttatcttttaaattttagaacaatttcacaggacatttcattcttcgagactttgaaaa	1176		
Db	1303	TAAATACTACTATTAAATTTGAATTATATNTATTAAGTTGTATTACATAGACCTTTACAA	1362		
Qy	1177	ataagaaatgaatttaagttaaaaaataaaaaa	1214		
Db	1363	ATGGTAAATAAAATATTATTTCGAATCAAAATATAAAATA	1400		


```

RESULT 14
US-08-213-419B-3/c
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: JII-003CNCP
; CURRENT APPLICATION NUMBER: US/08/213.419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5635)
US-08-213-419B-3

```

```
Query Match      3.7%; Score 44.4; DB 4; Length 837;
Best Local Similarity 54.2%; Pred. No. 0.11;
Matches 90; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 1049 taaatattataataataaaagatttctgcgatttatttagttatttaattagataaga 1108
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 TAAATAATGATAAAATATTAATTAAATAATTAATTAAATAATTAATAATTAATAATTAATA 529

Qy 1109 atttcatttcttattttaaatttaagaacatttacaggacatttacatttttcgagac 1168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 TGAATATAATTTTAAATAATATAAAATTAATAATAAGAAATTAAGTTAAATTAATTA 469

Qy 1169 ttgaaaaataaagaatgaatttaagttaaaaaataaaaaa 1214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 TTTTAATAATAATCTTATATAAAGATTAAATAATATAATCAACA 423
```

Search completed: August 27, 2002, 21:29:50
Job time: 3682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2002, 20:26:43 ; Search time 1866.38 seconds
(without alignments)
13611.813 Million cell updates/sec

Title: US-09-787-737-1
Perfect score: 1214
Sequence: 1 ctttagctcgtattatcat.....ttaaaaaaaaaaaaaaaaaa 1214

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
- 1: gb_ba.*
 - 2: gb_htg.*
 - 3: gb_in.*
 - 4: gb_em.*
 - 5: gb_ov.*
 - 6: gb_pat.*
 - 7: gb_ph.*
 - 8: gb_pl.*
 - 9: gb_pr.*
 - 10: gb_ro.*
 - 11: gb_sts.*
 - 12: gb_sy.*
 - 13: gb_un.*
 - 14: gb_vl.*
 - 15: em_ba.*
 - 16: em_fun.*
 - 17: em_hum.*
 - 18: em_in.*
 - 19: em_mu.*
 - 20: em_or.*
 - 21: em_ov.*
 - 22: em_pat.*
 - 23: em_ph.*
 - 24: em_pl.*
 - 25: em_ro.*
 - 26: em_sts.*
 - 27: em_un.*
 - 28: em_vl.*
 - 29: em_hum.*
 - 30: em_htg_hum.*
 - 31: em_htg_inv.*
 - 32: em_htg_other.*
 - 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	1214	100.0	1214	6	BD004607
2	1214	100.0	1214	6	BD004607
3	616.8	50.8	84510	8	AB006698
4	155.2	12.8	1518	6	BD004608
5	135.2	12.8	1518	6	BD004608
6	133.6	11.0	104596	8	U78721
7	108.2	8.9	175964	8	AF003760
8	93.4	7.7	91720	8	F2J10
9	87.4	7.2	81701	8	AB006706
10	73.4	6.0	89934	8	ATAC009327
11	69.8	5.7	1367	6	AX105291
12	64.4	5.3	95137	8	AC005560
13	63.2	5.2	113880	3	PFMAL3P4
14	62.6	5.2	53932	2	AC023371
15	62.4	5.1	74968	8	AB019230
16	60.8	5.0	506	6	AX105297
17	60.8	5.0	844	6	AX105299
18	60.8	5.0	6392	6	AX281395
19	60.8	5.0	6392	6	AX345586
20	60.8	5.0	204652	2	PFMAL13P6
21	60.2	5.0	966	8	AB058920
22	60.2	5.0	69188	8	AC007171
23	60	4.9	4481	3	AF466308
24	59.6	4.9	1136	8	PTRO10810
25	59.6	4.9	143879	8	AC069251
26	59.4	4.9	1037	8	AY048268
27	59.4	4.9	94315	8	ATF8D20
28	59.4	4.9	197859	8	ATCHRIV83
29	59	4.9	149259	9	AC009262
30	58.2	4.8	253305	3	PFMAL3P7
31	58	4.8	4601	3	DMU11584
32	58	4.8	19517	3	DMU37541
33	57.8	4.8	449	3	SFU06964
34	57.8	4.8	741	6	AX105301
35	57.8	4.8	164399	3	PFMAL3P6
36	57.8	4.8	224448	2	PFMAL4P4
37	57.8	4.8	321003	2	PFMAL4P3
38	57.4	4.7	93133	9	AC022140
39	57.2	4.7	110000	2	PFMAL13P2_1
40	57	4.7	199551	2	AC006281
41	56.8	4.7	1739	8	PCU42442
42	56.8	4.7	1826	8	AF196975
43	56.8	4.7	199551	2	AC006281
44	56.8	4.7	266371	2	AC020619
45	56.6	4.7	75803	8	AB016890

ALIGNMENTS

RESULT	1	BD004607	Homeobox	gene coding for protein participating in differentiation.	1214 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
DEFINITION	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
ACCESSION	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
VERSION	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
KEYWORDS	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
SOURCE	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
ORGANISM	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
REFERENCE	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
AUTHORS	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
TITLE	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
JOURNAL	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
COMMENT	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
OS	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
PN	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
PD	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
PF	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
PR	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002
PI	BD004607	Homeobox	BD004607	BD004607	1214 bp	DNA	linear	PAT 31-JAN-2002

Qy 61 caagcatgttcaaaatccaaaccttgcaacaataatcatcatcatcaacatgaatcgata 120
 Db 61 CAAGCATGTTCAAAATCCAAACCTTGCAACATATATCATCATCATCAACATGAATCGATA 120
 Qy 121 ctccattctacatgcactactcttaattgcaaccctatcaatctctctcttcctcagatcgga 180
 Db 121 CTCCATCTTACATGCACCTACTCTAAATGCAACCTATCATCTCTCTCTCTCTCAGATCGGA 180
 Qy 181 taccagatccttaaacccagagatgggaatccttaaacccgagcagatttaggatactcgaataca 240
 Db 181 TACCAGATCCTTAAACCCGAGATGGGAATCCTAAACCCGAGCAGATTAGGATATCGAATCAA 240
 Qy 241 tcttcaattccgggtactattaccacacccatagaggagatcacaagaatcggatccggc 300
 Db 241 TCTTCAATTCGGGTACTATTAAACCCACCTAGAGGAGATTCACAAGAATCCGGATCCGGC 300
 Qy 301 ttcaagaataatgggtcaaatcggtgacgcaacgtgttttactgggtttcaaacccggaat 360
 Db 301 TTCAAGAATAATGGTCAAAATCGGTGACGCAAAACGTTGTTTACTGGTTTCAAAACCGGAAT 360
 Qy 361 ctcgagcaaaacacaaagcttggtttcatcaacaaaagccctaaaatgtcacaagaaggaca 420
 Db 361 CTCGAGCAAAACACAAAGCTTGGTGTTCATCACAAGCCCTAAAATGTCAAAGAGAGGACA 420
 Qy 421 agacggttattccttagtactgacgtgatcatgttttgggtttgtttaaccaagaacccg 480
 Db 421 AGACGGTTATTCTTAGTACTGACGCTGATCATGTTGTTGGTTGTTTAACCAAGNAACCG 480
 Qy 481 gattatattccgggttcaaaaacaatgagttgggtggttaaccggaacccggtttctatttc 540
 Db 481 GATTATATCCGGTTCAAAACAATGAGTTGGTGGTAAACCGAACCAGCGCGGTTTCTATTTC 540
 Qy 541 cgggttcaataatgacccgagcgtcgtcaaatcagcgtttggtttggcgtatttggctgac 600
 Db 541 CGGTTCATAATGATCCGAGCGCTGCTCAATCAGCGTTGGTTGTTGGCGATTTGTTGTAC 600
 Qy 601 cgggtggaacggaagaagggtggcattctctaccgttaataaagcggttaatttggaga 660
 Db 601 CGGTGGTAACGGAAGAAGGATGCGATCTCTACCGTTTAATACGGCGGTTAATTTGGAGA 660
 Qy 661 ctaacgaaattttgtataaaattccggcgatcaattttatacggcgagatggaatggcg 720
 Db 661 CTAACGAAATTTTGATAAAATTCGGCGATCAATTTATACGGCGAGATGGAATGCG 720
 Qy 721 gtgaaattgttttctctcttctgactgtctcattcaaccatcaatcctcaagaataaac 780
 Db 721 GTGGAATTTGTTTCTCTCTCTTTGACTGTCTCATTAAACCATCAATCAATCTCAAGAAAAAC 780
 Qy 781 gagatgtaggattatccgggtggaagcgtcgagataaatgtttatccggtgagaatga 840
 Db 781 GAGATGTAGGATTTCCGGTGGTGAAGACGTCGGAGATAATGTTTATCCGGTGAGAATGA 840
 Qy 841 cgggtgttattaaacgagatgcctatcgagtagtgtctggattattcaacgttaaggcag 900
 Db 841 CGGTGTTTATTAAACGAGATGCCTATCGAAGTAGTGTCTGGATTATTCAACGTTAAGGCAG 900
 Qy 901 ctttcgaaacgagatgcggttttgatcaactcgttttggcagcctattcttcacagatgaat 960
 Db 901 CTTTCGGAACGAGTGCCTTTTGATCAACTCGTTTGGCCAGCCTATTCTTACAGATGAAT 960
 Qy 961 ttggtgttacttatcaacctctcaaaatggcgaactctattcttatttagaagatat 1020
 Db 961 TTGTTGTACTTATCAACCTCTCAAAAATGGCCGCAATCTATTCTTATTATTAGAGATAT 1020
 Qy 1021 tgaagaacaaatgttattggtctatggataaaattataataataaaagatttctgc 1080
 Db 1021 TGAAGAACAATGTTATGGTCTATGGATAAATATTATAATAATAAATAAAGATTTCTGC 1080
 Qy 1081 gatttatttagtatttaattagaagaatttcatttcttctttaaatttatgaaca 1140
 Db 1081 GATTATTATTAGTATTAAATGAGAAGAATTTTCATTTCTTCTTTTAAATTTATGAACA 1140

Qy 1141 atttacaggacatttaccatttccagactttgaaaaataaaagaatgaattaaagtaaaa 1200
 Db 1141 ATTTACAGCACATTTACATTTTCGAGACTTGGAAAAATAAAGAAATGAATTAAGTTAAAA 1200
 Qy 1201 aaaaaaaaaaaaaa 1214
 Db 1201 AAAAAAAAAAAAAA 1214
 RESULT 3
 AB006698
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MCL19.
 DEFINITION Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
 ACCESSION AB006698 BA000015
 VERSION AB006698.1 GI:2351063
 KEYWORDS
 SOURCE
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1 (sites)
 AUTHORS Kotani,H., Nakamura,Y., Sato,S., Kaneko,T., Asamizu,E., Miyajima,N. and Tabata,S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. II. Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned P1 clones
 JOURNAL DNA Res. 4 (4), 291-300 (1997)
 MEDLINE 98069011
 REFERENCE 2 (bases 1 to 84510)
 AUTHORS Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
 COMMENT Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MCL19
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://grembln1.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
 The 5' clone is K15122 and the 3' clone is MDE13.
 FEATURES
 Location/Qualifiers
 1..84510
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MCL19"
 /clone_lib="Mitsui P1"
 join(1437..1535,1857..1944,2030..2139,2260..2323,2912..2976,3055..3120,3321..3434)
 /note="gene_id:MCL19.1
 sp|Q38903"
 /codon_start=1
 /evidence=not_experimental
 /product="Rac-like gtp binding protein ARAC2"
 /protein_id="BAB08242.1"
 CDS

/db_xref="GI:9757717"
/translation="MSTARFKCVTVGDGAVGKTKMLISYTSNTPTDYVTVFDNFS
/ANVDGTVNLGDTWDTAGQDYNRLPLSYRGADVFLAFSLISAKSYENVTHKKWLP
ELKHVAPGPIVLVGTGLDRDDKQFLKDPHCAASITTAQGEELRKMIGAVRYLECSS
KTOQNVAVFDVAIVALRPPKAKKKTKRSRICEFL"
join(6597..6735,6885..7507)
/note="contains similarity to homeodomain transcription
factor"
gene_id:MCL19.2
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08243.1"
/db_xref="GI:9757718"
/translation="MSSSNKHNPMFKSKPCNNHHQHEIDTPSYMHSNCLSSSF
SSDRIPDKPRNPKPEQIRILEISGCTINPPREEIQRIRIQEYQIQIGDANVFY
WFONRKSRKHLRVHHSKPSKMKDKTVIPSTDADHCFGVNOETGLYPVQNNELV
TEPAGFLPVHNDPSAOSAFGFDVVPVVTGMAFSTVNGVNYLETNEFNDKIPA
INLYGDCNGGCNCPPLLVPLTINQSQGLLIISK"
join(9000..10120,10206..10271,10355..10499,10600..10819,
10909..11071,11182..11377,11457..11568)
/note="gene_id:MCL19.3"
/codon_start=1
/evidence=not_experimental
/product="CRN (crooked neck) protein"
/protein_id="BAB08244.1"
/db_xref="GI:9757719"
/translation="WTGDKAEVLPRTTRVKNKTPAPVOITAEQILREAREQAEIR
PKQKITDSTELDYLRRLRRKEFDQIRARNWIOVMKYAKWEESOMDYARARSWE
RALEGEYRNTLWVYAEFEMKNPVNARNVDSVTLPRVDQLWEKYIYMEELG
NVTGARQIFERNWNSPOQKAWLCFIKFLRNEIARSIYERVLCHPKVSATIRY
AKPEMRGGQVLAKEVYERAVDKLANDEEAELFVSFAEFEECKEYERAPFIYKFA
LDHIRKGAEELYKKFVAFEQYGDGEDAIVGKREYEDVSKNPLNDWFDY
VLEPSVGNKDIRIETRYAIRAIYPAQEKRFQWRYIYLWINYALYEETKDVKRYI
DYVRECLKLPHTKFSFAKILLAAEYERQLNLTGAROILNAGIKAPKVIKFIYI
EMELKLVKRCIKLYERELEWSPENYAWRYAEFISLAETERARAFELAIQPA
LWPELLWYITDFEISEGPEKTRALIERLDRTYKCHKWISFAKFEASAEHREDG
IKSAVIFORANTYITTPPELEERATLLEDWLNMETGFGELGVVVSQSLPKPLK
KKMTSRDGSYEYFDYLPPEESGTYTNLKILEAAYKWKLGEECY"
complement(join(111995..12297,12529..12809,12975..13130,
13269..13404,13629..13736,13846..14052))
/note="gene_id:MCL19.4"
/codon_start=1
/evidence=not_experimental
/product="myrosinase-binding protein-like; jasmonate
inducible protein-like"
/protein_id="BAB08245.1"
/db_xref="GI:9757720"
/translation="MSKVTIAYGSGNGKFDGYYEGTYDGVKLIYGEDFHHIVYL
KIQYKNGDVYVKEHGRAGTHITETEFVKCPDEYITSWGTCRNDHRYKSTGRT
FQOLTAADYIDINLAAMKPMIWNVTYSELQKTSHGRTSEQFRMPGTGSGNSWDD
GAYDRSLKLCVGEDAHCYSSVEFHYVYKGNDRITHCHGKDSKEHDKGFISSLFTKTS
NRSSEKFGPVGTQKQLEAKGFDKIVGFRGSSVNRINALGANAVVYVPPVKKLNKAK
GGVLGNWDDGTHDDARMTFKLYFNKEYITVSEGHYGRKLAAPNASAGMSFFTYG
MMLKPNTRNTTYQVLSHSPETYETGTSFKLEKDHKIYGVYKTEVSLNQIVYVK
IANA"
join(18260..18398,18630..18859)
/note="gene_id:MCL19.5
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08246.1"
/db_xref="GI:9757721"
/translation="MSSSNKSNKNDHDDHNDQGHNTIHTPSMHTPEISNISSPS
WPNRKYSRKPLVSYEMEL"
complement(join(20751..20785,20889..20954,21047..21132,
21235..21269,21369..21455,21539..21616,21701..21741,
22239..22305))
/note="gene_id:MCL19.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08247.1"

/db_xref="GI:9757722"
/translation="MGRGFKGKPTQRRFSSAADILACTSAAPRSFKOKEAYEED
VEESEESESEDEADYKKGAEAVIEVDNPNRVQKTLAKADLDASKTFLSRR
REELQORAHERYMRLEQEGKTEQARKDLRLALIRQOREAAKKEEKAARDAKV
EGRK"
complement(join(22934..23258,23634..23851))
/note="gene_id:MCL19.7"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="Rab-type small GTP-binding protein-like"
join(27158..27278,27480..27606,27840..27915,28196..28338,
28394..28464,28552..28573,28658..28814)
/note="gene_id:MCL19.8
pir||T27174
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08248.1"
/db_xref="GI:9757723"
/translation="MLGHSSQDNFYASDEVSTMAFLLLAFSTSRVSSPDCLEEREI
KTMKRVRENEEKALWFRQFDLVKCLLSKCLNGLHLSPPDRHLLRNHLN
NTHPLVLAKQWERGNOEQLLRSEWISLTSQSLVLSAITGLVSNASCWNHNDMKH
LGKACRICESESFTIITALEADIYSEWIDECERVATAEDDVVQVEEEVEE
EEEEDEDDHVSVKRYNF"
complement(join(29073..29934,30037..30602,30797..31014,
31113..31227))
/note="gene_id:MCL19.9"
/codon_start=1
/evidence=not_experimental
/product="peptide transporter"
/protein_id="BAB08249.1"
/db_xref="GI:9757724"
/translation="MTVEEVGDYTKDGTVDLGRNVRRSOTGRMKACSFVVVVEFE
RWAYGISNLVIYMTTKHQGVKSSNNVTNWVGTSLTPILGAVVADAHFGRYITF
VYSAYILGLMALLTILSVSLGPKPCSTANVENCEKASVQLAVFAGLYTLAIGT
GGTKPNISTIGADQDFDPDKIKHHSFFNMWFSIFFGTFTATLVLYVQNVGHA
IYGSULTGLAFSIFILFTGLRLYRHKLPMSGPTTKMARVIVASLRKAREPMSSDSTR
FYELPMETASKRAEPIHSTSLRFLNRASLKTGTHKWLCTITVEETKQMLKLP
VLFVTPVSMIAQIMTLTKOQTLDRRLTNNSIPSPASLSGFTVLSMLISIVLYDR
VFVKMRKLTNGRITLQRMGIGMILHILMIASITERVRLKVAABGLHTQAV
PIPLSIFTLPOVLMGLADAFIEAKLEFFYDOAPESMKSIGTSTVTSMAVGYEMS
SILLSVSOITKKQGRGIONNLSRLDNYMFFAVLNLNLFILVIRFVEYRAD
VTQSANVEQEPNWDNYNE"
complement(join(32738..33188,33279..33677,33771..34336,
35792..36009,36633..36747))
/note="gene_id:MCL19.10"
/codon_start=1
/evidence=not_experimental
/product="peptide transporter"
/protein_id="BAB08250.1"
/db_xref="GI:9757725"
/translation="MTVEEVGDYTKDGTVDLQGNPVRRSIRGRMKACSFVVVVEFE
RWAYGISNLVIYMTTKHQGVKSSNNVTNWVGTSLTPILGAVVGDALLGRYITF
VYSAYISGSMVLTISVTPGKPEECSTTNVENCEKASVQLAVFAGLYTLAIGT
GGTKPNISTIGADQDFDPKPKTSFFNMWFSIFFGTFTATLVLYVQNVGHT
LGYSLTGLAISITIFLTPFPRHKLPGSPPTTKMARVIVASFRKANAPMTHDITS
FHLEPLSEYERKGAFFIHTPSLRLDRASLKTGTHKWLCTITVEETKQMLRMP
VLFITFVPSMLAQINTLVKQGTLDKRVTSFSIPSPASLSGFTVLSMLISIVLYDR
VFVKTRKFTGNPRGTLQRMGIGIHLIMIVASVTERYRLKVAADHGLHTQAV
KLPTIFALLPOFLMGMADFSLEVAKLEFFYDOAPESMKSIGTSTVTSLAIGNFMS
SELLSTSEITTKRGRGWIILNLSRLDYYLFFAVLNLNLFVLFVVKFYVYRAE
VTDSDVQKVEEMKETE"
join(39209..39382,39563..40135)
/note="emb|CAB89385.1"

Query Match 50.8%; Score 616.8; DB 8; Length 84510;
Best local similarity 83.7%; Pred. No. 1.4e-109;
Matches 777; Conservative 0; Mismatches 2; Indels 149; Gaps 1;
Qy 1 ctttagctctcgattatcatcattacacattcatctctctctcaacaaaaattggc 60
|||||
Db 6562 CTTTAGCTCTCGATTATCATCATACCATCATCATCTCTCTCTCAACAAAATTTGGC 6621

QY 61 caagcatgttcaaatccaaaccttgcaacaataatcatcatcatcaaatgaaatcgata 120
|||||
Db 6622 CAAGCATGTTCAAAATCCAAACCTTGCAACAATAATCATCATCATCAATGAATCGATA 6681
QY 121 ctccactcttaactgactactactataatgcaacctatcatctctcttcttccct----- 171
|||||
Db 6682 CTCCATCTTACATGCACTACTCTAATGCAACCTATCATCTTCTCTTCTCCCTCAGGTACTC 6741
QY 172 ----- 171
Db 6742 ACAAAATTTACATTTCTCTGTTAAATTTATAAAAGTATATTGTTAAATTTACTTGAT 6801
QY 172 ----- 171
Db 6802 TATACAAACAACTATATGAACGATGTTATGTATATATGTTAAGACATGCATGCATGATG 6861
QY 172 -----cagatcgataccagatcctcaaatcggtgacgcgcaaa 211
|||||
Db 6862 ATATAAACCGCTCTATGAACAGATCGGATACAGATCTTAACCGAGATGGAATCCTAA 6921
QY 212 accgagcagattagatgactcgaatcaatcttcaattccggtactattaaacccactag 271
|||||
Db 6922 ACCGAGCAGATTAGGATACTCGAATCAATCTTCAATTTCCGGTACTATTAAACCCACCTAG 6981
QY 272 agagagagattcaagaatcccggtccggttcaagaataatggtcaaatcggtgacgcgcaaa 331
|||||
Db 6982 AGAGGAGATTCAAGAATCCGGATCCGGCTTCAAGAATATGTCATAATCGGTGACGCAAA 7041
QY 332 cgtgttctactggtttcaaaacccggaataatctcgagcaaaacacaaagcttcgtgttcatca 391
Db 7042 CGTGTGTTTACTGTTTCAAAACCGGAAATCTCGAGCAAAACAAAGCTTCGTGTTCATCA 7101
QY 392 caaagcctcaaatgtcaagaagacaagcgtttattctactgactgacgcgtgatca 451
|||||
Db 7102 CAAAGCCCTAAATGTCAAAGAGGACAAGACGGTTATTCTTAGTACTGACGCTGATCA 7161
QY 452 ttgttttggtttgttaaccagaacccggtattatctcgggttcaaaacaaatgagttggt 511
|||||
Db 7162 TTGTTTGTGTTTGTAAACGAACCGGATTATATCCGGTTCAAACAATGAGTTGGT 7221
QY 512 ggtaaccgaacccggtgtttctctattcttcgggttcaatgacgcggtgcctcaatc 571
Db 7222 GGTAAACGAACCGGCGGTTTCTATTTCGGGTTTCATATGATCGAGCGGTGCTCAATC 7281
QY 572 agcgtttggtttggtggtttgttaccggtggttaacgaagaaggtatggtcattctc 631
Db 7282 ACCGTTTGGTTTGGCGAATTTGTTGTACCGGTGTTACGGAAGGAGGTGGCAATTC 7341
QY 632 tacggttaataacggcggttaatttggagactaaacgaaaatttttgataaaattccgcgcat 691
|||||
Db 7342 TACCGTTAATAACGGCGTTAATTTGGAGACTAAACGAAAATTTTGATAAAATTCGCGCAT 7401
QY 692 caatttatacggcgagatgaaatggcggtggaattgttttctctctcttgaactgttcc 751
Db 7402 CAATTTATACGGCGGAGATGAAATGCGGGTGGAATTTGTTTCTCTCTCTTGAAGTTCC 7461
QY 752 attaacatcaatcaatctcaagaaaa 779
|||||
Db 7462 ATTAACCATCAATCAATCTCAAGGTAAA 7489
RESULT 4
LOCUS BD004608 1518 bp DNA linear PAT 31-JAN-2002
DEFINITION Homeobox gene coding for protein participating in differentiation.
ACCESSION BD004608
VERSION BD004608.1 GI:18632569
KEYWORDS JP 03074789-T/2.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1518)
Kakimoto,T.
Homeobox gene coding for protein participating in differentiation
Patent: JP 03074789-T/2 26-JAN-2001;
SUNTORY LTD,TATSUO KAKIMOTO
OS Arabidopsis thaliana (thale cress)
PN JP 03074789-T/2
PD 26-JAN-2001
PF 21-JUL-2000 JP 2000004904
PR 22-JUL-1999 JP 99P 207995
PI TATSUO KAKIMOTO
PC C12N15/29,C12N1/21,C07K14/415,C12P21/02,C12N5/10,A01H5/00// PC
(C12N15/29,C12R1:91),(C12N1/21,C12R1:01),(C12P21/02,C12R1:01) CC
FH Key Location/Qualifiers
FT CDS (152)..(1285).
FEATURES
source 1..1518
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
BASE COUNT 425 a 389 c 279 g 425 t
ORIGIN
Query Match 12.8%; Score 155.2; DB 6; Length 1518;
Best Local Similarity 63.3%; Pred. No. 2,6e-20;
Matches 238; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 34 tcatgtctctctcaacaaaaatggccaagcatgttcaaatccaaaccccttgcaacaata 93
Db 150 TTATGGCTTCTTCGAATAGACACTGGCCAAAGCATGTTCAAGTCCAAACCTCATCCCATC 209
QY 94 atcatcatcatcaacatgaaatcgatactctccatctcttaccatgcactactacttaattgcaaac 153
Db 210 AATGCAACATGACATCAACT 269
QY 154 tatcatctctcttctctcagatccgataccagatcctaaacccagatggaatcctcaaac 213
Db 270 TCTCTTCAGGATGTCAGTGGAGAGAGTCCAGAGCAAAACCAAGATGGATCCAAAGC 329
QY 214 cggagcagataggatctcgaatcgaattcttcaattctccggtactattaaacccacctagag 273
Db 330 CAGAGCAGATTCCGATCTTGAAGCAATCTTTAACTCCGGGATGTTGAATCTCTCCAAGAG 389
QY 274 aggaattcaagaatccggtatccggttcaagaatatggtcaaatcgggtgacgcaaacg 333
Db 390 AGGAGATCAGGAGGATTAGGGCTCAGCTTCAAGATACGGCCAAGTCGGTGTGCTTAACG 449
QY 334 tgttttactggtttcaaacccggaatctcgagcaaacacacaaagcttcgtgttcatcaca 393
Db 450 TCTTCTACTGTTCCAAACACCGTAAGTCCCGTAGTAAACACAAACTCCGCCCTCTCCACA 509
QY 394 aaagccctaaatgtc 409
Db 510 ACCACTCCAAACACTC 525
RESULT 5
LOCUS E52178 1518 bp DNA linear PAT 31-JAN-2002
DEFINITION Homeobox gene encoding protein participating in differentiation.
ACCESSION E52178
VERSION E52178.1 GI:18629641
KEYWORDS JP 2001029081-A/2.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1518)
Kakimoto,T.
Homeobox gene encoding protein participating in differentiation

For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=MVA3
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail (<http://informatics.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>), NetGene2 (S. M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremmlin2001.iastate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
 The 5' clone is K10A8 and the 3' clone is MP17.

FEATURES

Source

Location/Qualifiers

1..81701
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MVA3"
 /clone_lib="Mitsui p1"
 complement(1..13)
 /note="CDS is reported in Acc# AP002488"
 gene_id:K10A8.130
 /number=1
 /evidence=not_experimental
 /product="glycine/proline-rich protein"
 complement(join(1028..1252,1346..1444,1554..1638,1732..1781,1904..1983,2360..2435,2523..2579,2795..2965))
 /note="gblAB08476.1"
 gene_id:MVA3.1
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09565.1"
 /db_xref="GI:9759043"
 /translation="MREHKIPHCNFRFNAFNAVSASSASTSVCVDFKDLRSNELVEL
 EYALNHNKISEEVGHVRIHQVNPSSFSAPAPVPVWDEYKPSLPMDVIGSG
 SGFLQWANKNYSYGLERQKLVKRNANFVNLGLSNVHFIFANASVFEHLI
 SYPGPLEVYSILCPDPHFKRQKRVQKPLVNSILQNLKPGGKIFVQSDVLDVAQ
 DMRDQDEESNVLQHMDTVDTEGWLTPNMGIRTERIHAEPGARIVRYRQKRL
 LPT"

exon

CDS

join(3332..3448,3674..3846,3937..3989,4101..4215,4308..4430,4502..4850)
 /note="gene_id:MVA3.2"
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09566.1"
 /db_xref="GI:9759044"
 /translation="MASAMVTSURLPNFNSAATSPSSSSAGNVKVRPAVILPGLN
 TGYKLEVTLGEGYPAVVAASRLDWFNAAGLVDPAVWGTGLRPRVLDWLNRI
 DDAVREANLAQOGLGHSAGWLARYMEYGNDSISLLTLGLTGLPPLPPGLP
 GVDTPRGLLYVEENCACAVYTPELKYVICIAGRIYRGARLVNADADIDSVTGID
 SGGISELAIASNKKSGTFRARFVGQYQVCGRADWGDGVVPEVSAHLEGALNVSE
 DGYHSPVGSDDTERWYGSVPVVKWNIHLLLE"
 join(5249..5718,5796..6894,6993..7292,7385..9403)
 /note="gene_id:MVA3.3"
 /codon_start=1
 /evidence=not_experimental
 /product="disease resistance protein-like"
 /protein_id="BAB09567.1"
 /db_xref="GI:9759045"
 /translation="MASLPSSSSSSSSTVMKTDVFSFRGEDVRKTFVSHLCEFFDRM
 GKAFRRDLDQKGSISPELIDAIKGSFPAIVVSRNAAASWCLDELKIMECKND
 TIYPIFEVDPSVRRQSGFEDVESHSDEKVGKWEALKKLAISGDSNRWDD
 SKLIKIKVDISKLVSSTWDDSKGLIGNSHMDFLQSMISITVDKVRMLGIMGVGV

CDS

join(3332..3448,3674..3846,3937..3989,4101..4215,4308..4430,4502..4850)
 /note="gene_id:MVA3.2"
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09566.1"
 /db_xref="GI:9759044"
 /translation="MASAMVTSURLPNFNSAATSPSSSSAGNVKVRPAVILPGLN
 TGYKLEVTLGEGYPAVVAASRLDWFNAAGLVDPAVWGTGLRPRVLDWLNRI
 DDAVREANLAQOGLGHSAGWLARYMEYGNDSISLLTLGLTGLPPLPPGLP
 GVDTPRGLLYVEENCACAVYTPELKYVICIAGRIYRGARLVNADADIDSVTGID
 SGGISELAIASNKKSGTFRARFVGQYQVCGRADWGDGVVPEVSAHLEGALNVSE
 DGYHSPVGSDDTERWYGSVPVVKWNIHLLLE"
 join(5249..5718,5796..6894,6993..7292,7385..9403)
 /note="gene_id:MVA3.3"
 /codon_start=1
 /evidence=not_experimental
 /product="disease resistance protein-like"
 /protein_id="BAB09567.1"
 /db_xref="GI:9759045"
 /translation="MASLPSSSSSSSSTVMKTDVFSFRGEDVRKTFVSHLCEFFDRM
 GKAFRRDLDQKGSISPELIDAIKGSFPAIVVSRNAAASWCLDELKIMECKND
 TIYPIFEVDPSVRRQSGFEDVESHSDEKVGKWEALKKLAISGDSNRWDD
 SKLIKIKVDISKLVSSTWDDSKGLIGNSHMDFLQSMISITVDKVRMLGIMGVGV

CDS

join(5249..5718,5796..6894,6993..7292,7385..9403)
 /note="gene_id:MVA3.3"
 /codon_start=1
 /evidence=not_experimental
 /product="disease resistance protein-like"
 /protein_id="BAB09567.1"
 /db_xref="GI:9759045"
 /translation="MASLPSSSSSSSSTVMKTDVFSFRGEDVRKTFVSHLCEFFDRM
 GKAFRRDLDQKGSISPELIDAIKGSFPAIVVSRNAAASWCLDELKIMECKND
 TIYPIFEVDPSVRRQSGFEDVESHSDEKVGKWEALKKLAISGDSNRWDD
 SKLIKIKVDISKLVSSTWDDSKGLIGNSHMDFLQSMISITVDKVRMLGIMGVGV

GTATTAKLYNQLSGQFVHCWENVCNRYGVRRLQVLEFLCRMFQBRDXANSSV
 SCNIIKFRHKMVFILDDVDVRSQELNKLAKETGFGPGRSRIIVGLTKRHLHLSHG
 INLVYKCLPKKEALQFCNAFREELIILPHGFELSQAQVNAVAGTLALVLSGFS
 LYRRSDVEESTLARLTYPKSDIMEVLRVSYDGLDEQEKAFILYISCFYNKQVQYV
 RLKLDICGVAAEIGITILTKSLIVESNGCVKIHDLQEQMRELVRQQAQVNNQORLL
 LWDPECHLSENSTQLVEGSLNLSEISEVAFSDRAFEGLSNLKLNFYDLSFDG
 ETRVHLPLNGLSYLPKRLYRMDGYPLKTPMPSRFFFEFLVELCMNSNLEKLDWGTOP
 LRNLKMDLSRCXYLVEVPLDSKATNLEENLNSYCSLSVLETPSINKLGLSCFYLTN
 CQLKDPICITLKSLETVMGSCSLKHLKHFPEISWNRRLYLSSTKIEELPSSISLS
 CLVLKMSQCORLRTPSYGLHLVLSLNLDCGRRLNPDQLNLTLETLEVSQC
 LANVEPRVSTSEVLRISETSEETIPARCINLSQSLDISENKRLASLPVISELR
 SLEKLKSGSVLESPLEICQTMSCLRMFDLDDTSIKELPENIGNLVALELOASRT
 NISGNLNLLELDLSDNNEFFIPASIKRLTRLNRLNLCQALQALDELPRGLLIY
 ISHSTLSISGCFNOYCLKRLKSNYKLDQAAQILIHRLKLESAKPEHSYFPQSD
 IPTCFNHQVMPSLNQLPQSSSSDILGFSACIMIGVDGQYPMNLLKIHCSKILDA
 DACELVVMVAVPDKAFTNMYFGSDHLLLSFTCTCTSMAYSEALFEFSEVTEGDS
 FSPLEGVKCAVHLISLKMOMQEFNSDKIQSSDLDLPKAFDDARILKRAYETEFL
 HKQLNPKRIKFLPVLARERHSLKHQRNVRK"
 complement(join(9734..9801,9890..9950,10033..10788,10901..10987,11089..11114,11428..11787))
 /note="contains similarity to polycomb protein"
 gene_id:MVA3.4"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09568.1"
 /db_xref="GI:9759046"
 /translation="MKGASGAVKKPQVLNEAGEAETAETVETGESRKISGDGFGSD
 GGGGGGGESILREIGDDPTEDGDEDEDEDDGDEDEDEGEDEGEDEGEDEGEDE
 LDGFYEIAIRBRKVRKGVOLIKRWGPETANTWEPLENQSADVADAFEGSLK
 PKGPKRRKRYAGPHSQKKKORLSTSHDATEKSDSSTSLNNSLIPDIPDLDSG
 SLLNRDVEAKNAYVSNVEANSVGMAQVRLIDNEKSTDETLNELRGFVNNNGA
 GCSQGGIGSEGDVNRPLGLLVYPKELDKNRFIGAKRRKSGVRFKQDGSSTNNH
 TSDONLTPDLTLDISFGRIARMNEYPGVMENCLNSQTKTEELDTILKPMSTP
 ASVDNVQVLYTFLALRSDGKEALVDNRKALHNPHLLIEFYEQHLKYRNP"
 complement(join(13436..13537,13635..13753,13858..14096,14210..14266,14341..14427,14564..15105,15399..15665))
 /note="gblAAF03470.1"
 gene_id:MVA3.5
 strong similarity to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09569.1"
 /db_xref="GI:9759047"
 /translation="MSGGGEEMERLNGSETQRRSLYLKRNKINSEVRKMRALP
 STLRVMSFCVVVAQAFITGSSETGLAAYLIQSTFIRFIYIMAGMSATETLCQO
 AYGAQYHMGYILQRSWIVDTFIATLFPFIVLAGILRLLLGNVVISVDEIYPM
 VYPLYSIVFTMTQMQLQAKNAIIGLSTLALVLDIAATWCVSMVGMGIGHALL
 GLNLSWSVAIEFVYVFGGCPHTWTGFTAAFLDLIPMLKLSISSGMLCLEYWYM
 SIIVLSGYAKDANIALSAFSCQYLYSWEMNICFGLMGAACVVRVANELKGADAVR
 FSIKVLVSAVIGVICSALCFLAGGQISYLFSDSQVSDAVADLSIVLSISILFNLI
 QPILSGVAIGAGHQSVALVNLASYAIGVPLGLVLYVFNFGIKGLWSGMLAGVGIG
 TLILCYVYKTDWELEVIYF"
 complement(join(21810..22022,22560..23075,23638..23889))
 /note="emb|CA04381.1"
 gene_id:MVA3.6"
 /codon_start=1
 /evidence=not_experimental
 /product="GrpE protein"
 /protein_id="BAB09570.1"
 /db_xref="GI:9759048"
 /translation="MAGLLKTPSLHLPTLLHAPSPVFPKPCVSAFAGRNVSLSRR
 ASLRVSSGYPRLNLNLPFASGEAETETEVSENEPQETDQAGVVENENASAECS
 EABEEAAVITALLKSYKALADNNEKRIAEIESLASIETEDERFLLDKQVLSNELS
 VERDLIRISADFDFNRKTERLNLVNSAQGVENLNLAVLDNFRKASQIKVETE
 GEKVTNSQSYIQFVEIIGSLGVIHETVQKQFQDPLMHEANREDSAEYEEGIVLE
 EYRKFLGGERLLRPSMVKSAGPGPEKLEAEGEATEAQSAAEESSSS"
 join(28184..28637,28716..29062,29156..29352,29570..29903)
 /note="gene_id:MVA3.7"
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental

1294 aa, predicted using hexExon"
/codon_start=1
/product="hypothetical protein, PFC0435w"
/protein_id="CAA15616.1"
/db_xref="GI:3764023"
/db_xref="SPTREMBL:O77361"
/translacion="MKVGIIFFCLFFVFWLGGACNVKRIFFKNIKKTKFTLLNEPDN
DLGSFENLFTLLFDLDKNTLFDLSLNLLENLNYSSIFRLAVDVPYKTKNEEDN
KNRYIFLGTSFISRIHPLFEYELKLNKYIYAGNIYEKGNVDIRGLIEYKREIEK
KLEKOLKNKRLNNNNNNNSKFSKGDGNEEDFNNDLNPSSDKLYNNDDIDVHE
LLEEITTKERKFLDODDDNDNDKYILKTDENVKRYGFFIGYGFNDIDPSVHHYNE
DKNFFPSLNSGILDTLLKNYIEFVYKYOVNLFKDYKNTEKQYKIKENLV
RLTHSEGENLEONEIHLNDPNNEFIYKYOVNLFKDYKNTEKQYKIKENLV
RHETSSEGENLENNRNTKHNNNDNNNNSEDAIALLSYFNVPYPISTCMCYSTR
SKHESLDYKYMHNLENDIKLHYIKETEEHFNSIEBKMKLNRLNKKYDILLSE
HENLVTHDKYIKTSINTEERKPHIKNTYDNKNTQIIFNTKNTLNKLNKLNKTFG
FYNNSLQNALENDNDIDDIYMSDKESQYDNLNFKSVTSKGLCEKLNKHMIIYY
EYVMKNSKKYFFIADDDYFNKYLIDVTNLNTCSHSKKYMDKYTKSDYFVE
NEALFLQNFPKTKFLFLYSLKQDTFAKTIQLKKYDVPKYCOGGILSKKHKNDSDDD
HDHVGKQNDSTNHODIEKNOVNVINNNNNNNAKSIPIVLRGRYSNTFTNS
NEFYDYLTCGAGILINDETAKRIYECKECTCPSTNSSMDMIFGKAKELGILAINP
EGYFONSPLYNNKYINTLPIYHRLNKRRTKESRDMYFNLYNVRNDKEONKDI
YVDLDRHNNNDNPHYFFYVMYDEKKNVYTKIEHNADMMNSKKNKSNPKQKLN
QGKNVNDNDNNDENKGVNKGDNVKGDEYMKGDNVKGDNVKGDNVKGDNVKGDE
NIRGDDNNVNDNDNDIDINMYESYDDVMERNKKGCTGKEKKDDRNHNKKEATDV
KIKSVPTNNIKNEDTTKYVKMNEKIYNRMQESGKQLFDINKFPKKEIEGHPYFO
KIKKNEKAKKEKNQAKQKDYTNVYFHTSNMOCNFNOQKMGVONQONENDEDFE
DORPEIEDAINPDYIEYMNLSNFEDEGEPEYDEYDDYDDFYNTIADKLKINDQNK
HLFEQIDIAQPPVNFONDQNSNTFFDFTDEL"
6245. .6250
/note="potential splice donor sequence at end of PFC0435w
may suggest splicing"
complement(7159. .7164)
/note="potential splice donor sequence at end of PFC0440c
may suggest splicing"
complement(7161. .13970)
/gene="PFC0440c, MAL3P4.25"
complement(7161. .13970)
/gene="PFC0440c, MAL3P4.25"
/note="PFC0440c (MAL3P4.25), putative helicase, len: 2270
aa; Similarity to helicases. C.elegans helicase
(WP:F52B5.3) BLAST Score: 290, sum P(4) = 5.9e-31; 34% in
199 aa overlap, predicted using hexExon, contains match to
PF00271 helicase_C, Helicases Conserved C-terminal domain
Score 25.70"
/codon_start=1
/product="putative helicase"
/protein_id="CAA15615.1"
/db_xref="GI:3764022"
/db_xref="SPTREMBL:O77360"
/translacion="MNTFKINKKRRTVYESYNTIANKKRYMCNDKMDNDNNIIN
EKYFNSSIKNTGNNYKNDIDKSYDMCERBSMDRDENIYNNYKKNRNNYSNK
NNYSNNIHNDRANTYKGLDGRKQYLSNKHESKNDNSYNLKKYNFIPKIKYN
NNNNNNNNNDNRQNDIINLYFYDDKNEKKSNTHTSYGNFNNEFTNPKQNTM
NSNEHNTQDCHPTIYHNNNNNDGRKIDISLSYGNKISDNDLSKENTSYINQY
NSGGAHRVYIAQHNHNNHNNHNNNNYHNYNAKQYDKYKNGDSKFSAYNAE
QHDMDKYTDIHKNNKHSTKGDATLSFGOKVFGSEKKNQMKKRPKPKNKRGS
SINRPGMHHNNYNNNNNNYNNYNNHFCDDAERNNKNEKCEINSNELYEKDG
MPYINANEENLKKMYKDKKEKIKILLTDENVNQQNSIEKLSYKSRNEIEM
TERNDVTFINGETSGKSTCVPKFLEENIRENKKIINIVTEPRRIACIALSKLTSEL
TNEKLGQKIGYRISGESYDSEKTVITYITIGYLFKLFLHKKNYKRFTHVIDEIH
RSLIDVLVLFKILYHLNKKQDEOMFLIIMTASQNLFFSYEPENIKMGSIFIGT
KIFSIDFYTDIINYTRYGRSNIHDEKINDEKINDEKINDEKINDEKINDEKINDEK
DKKSCDDKSCDGNNEFKESVIDFILRRKCNKINLSKSEMLLYIKSBDKNIHM
FNNNNKYCNKNDKLDVDELIPANFNSNISCLELYVNLGKDSVIIFLSQMDI
TDMYHQUSMINNDNPNNSNINNGNDANDANNHNNHNNHNNHNNHNNHNNHNNH
HRTDIKTHIHLHSLCYDNTIHLKLNHNDTINIFLSSNIAESSITVNLVIRIDFCIQ
KNYENDKKAWILVKKWINKSMEQRKGRGTCGICIRMIKSNFLNLRDHIKISE
IYHSLHLIYLYILKSNVNLGNLKNREVVHCDTNYVNVKVDKQNGICKNEISH
SHAENKKLTIYDLVLSIERKSIKSTRYELEKRAVAKIKDKLVISIIIGQIMIRP
NLSINLCRLLYLVLLDVTFTDITILIGILANTDFFPINLYSSKNYSYAVSLGICR
QKAYFDGNTYSEPIMLNVLFWLCVYLLYVQGLKKNKFKRKLKTYVMTCISIMNK
RNHVNSKKLLCVINSVDNCKKMLKMLNKNNSNAYKSCVYLLYLLRGGSDINYNNGNS

NNNDNGSEIINADGITPNNTHMNEINGMNVNVVTKYCLFDYSNQNIYLKFLSLSF
TPLIHGSPNLPRLNLEGKKSKKLMNLNFMIDNKLDKVKNCIYFSGVYIPDINILK
RAICIMCPISFDIYVSKNIYYVYFHNHNNYHNNHNNILKGGDSYNNKTCIDGAD
MSVNSYLNLTQEGHDTDGTISFGVSPLMNNHNNHNNHNNHNNHNNHNNHNNHNN
NSQNNYMMKMNISNYSESLNLTSGNKYPIIIEKIRNKLILYKNNHNNHNNHNN
NNIILNQPVNAIVPYSNKYDEINNASICTNINMFSGKWIETMPLFNPKNRKCSEY
SEYNHHELYELKPKHLFLVKWMLNTDNYKIKKNQKQEHVMTNALLNNSLGNKVS
KNDIHRNIDKSLNDOEDQYNGYVYDDYDDYDDYDDYDDYDDYDDYDDYDDYDDY
NNDDYLSNVKVESDNNHNTQNPDPGEYEGNDSMCSNNKSKQKKMKCNLPRSV
LGFUSMCPFNEDTARNIYDQTADIVACSSIDYSYNTDITYWNVITVLPKNTYFLT
FLSSIPYDHNVIIOITRNLYSDILSIKIFDSKEIILKMMKMKHKNKNTNTYDTN
STINNPNTIHDLLRVNLYRVMSKLLLSLSPNRQIEQNTOPKLSKSKRKSIV
INKDGNPNDRVDANDNRYMKNNDKVNQDNQSNQDNQKOHFDDDEYFSDSD
IMSDIENILKENLYLKOIIVNTEENMDYINNYKSYMNDMDIDERNFLQIYLEDE
HFNLNKKWNNKGEILMNTYDLYLSLKNENDYNOYNNNNNNNNNNNNNNSSYVKN
KEKALNSLEAVKSSYLLNNNFSVQVNTNAYERTNSDLEFLFPQINISIPKEANKLR
SMYNEVAPTLDKFRRTNRQV"
complement(10956. .11255)
/gene="PFC0440c, MAL3P4.25"
/note="Match to PF00271 helicase_C, Helicases conserved
C-terminal domain Score 25.70"
15740
/note="variable sequence, 9 or 10 Ts"
complement(join(16039. .16667,16813. .16891))
/gene="PFC0441c"
complement(join(16039. .16667,16813. .16891))
/gene="PFC0441c"
/note="PFC0441c, hypothetical protein, len: 235 aa,
revised: new gene prediction"
/codon_start=1
/product="hypothetical protein, PFC0441c"
/protein_id="CAB89483.1"
/db_xref="GI:7672213"
/translacion="MDININMEYQAIVLPHMFMNASKLVSLDYVAVVAVPENNEYD
NDWMLTLRLFCLYPHCYVDKDIIRKILICLYNINNVDMNMMSLINSLLDNDIKSV
IYIYELIKNCQYKKLWNCINNINNNYNNCDYSYLKNDKNFCINRILYKILTSISFE
SYILKNNSELYNIQDNTQLEQFLNENKNTIRKINHKGDEQICNGTETVQNKKNIN
TYFTEDNIGSYITKLNH"
complement(16668. .16677)
/gene="PFC0441c"
/note="PFC0441c"
complement(16807. .16812)
/gene="PFC0441c"
/note="PFC0441c"
complement(join(18192. .18242,18754. .19140)
/gene="PFC0445w, MAL3P4.24"
join(18192. .18242,18754. .19140)
/gene="PFC0445w, MAL3P4.24"
/note="PFC0445w (MAL3P4.24), conserved hypothetical
protein, len: 145 aa, revised: extended upstream and
added exon, GC plots suggest more, similarity: eg to human
cDNA AAF29135 HSPC172 (219 aa), fasta scores, opt: 257,
E(): 6.9e-11 (40.0% identity in 115 aa overlap)"
/codon_start=1
/product="hypothetical protein, PFC0445w"
/protein_id="CAA15613.2"
/db_xref="GI:7672214"
/db_xref="SPTREMBL:O77358"
/translacion="MSLYVNNQHGTLVYQKHFSFEIKLSNEEIRLASMHLGISTIS
EKNVYSLYEKKTIIQSLKEKKEITIEGDFKIQCYDTLTGKIFIVHKDDLNTEM
TYLNRVVELYSDIILKNPFYDIDPIRSVAFNEQIEKLSNIS"
19528. .19857
/gene="PFC0450w, MAL3P4.23"
19528. .19857
/gene="PFC0450w (MAL3P4.24), Hypothetical protein, len:
109 aa, contains possible signal sequence, predicted using
hexExon"
/codon_start=1
/product="hypothetical protein, PFC0450w"
/protein_id="CAA15614.1"

```

/db_xref="GI:3764021"
/db_xref="SPTREMBL:O77359"
/translation="MSMFLNILILDAASVAFLLITFLMNLNEESLELSQAHRENGK

Query Match          5.2%; Score 63.2; DB 3; Length 113880;
Best Local Similarity 55.5%; Pred. No. 0.013;
Matches 122; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 995 aatctattattctatttagaagattgaaagcaaatgttggtgctatgataaata 1054
      ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103570 AATGGAATAACATACGTAATATATATATATATATATATATATATATATATA 103511

QY 1055 ttaataataataaagattctgcgatttatttagttatttaattagataagaattca 1114
      | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103510 TATATATATATATGATGATATTTTTTTTTTTTATGAACAATTCAGGAGAAAAAATTTT 103451

QY 1115 ttcttattcttcaatttgaacaaattcacaggacatttcatttcgagactttgaa 1174
      ||| ||| ||||| ||| ||| ||| ||| ||| ||||| ||||| ||| |||
Db 103450 TTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTTATGGCTATATTTTCAATTTAGCTGAAAAAA 103391

QY 1175 aaataagaatgaatttaagttataaaataaaaaaa 1214
      ||| ||| ||| ||| ||| ||||| ||||| ||||| ||| ||| |||
Db 103390 AAAAAAAATAAAAAAATAAAAAAATAAAAAATAA 103351

RESULT 14
AC023371/c
LOCUS Homo sapiens clone RP11-21D18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION Homo sapiens clone RP11-21D18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023371
VERSION AC023371.2 GI:9123990
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 53932)
Birren,B., Linton,L., Nusbaum,C., Lander,E., E. Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,E., Boguslavsky,L.,
Bouknight,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Collangelo,M., Collins,S., Collumore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,B., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,B., Olivari,T.M.,
Peterson,K., Pierrre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6970502.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

```

```

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3985
Center clone name: 21_D_18
-----
* NOTE: This record contains 55 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 869: contig of 869 bp in length
* 870 969: gap of 100 bp
* 970 1815: contig of 846 bp in length
* 1816 1915: gap of 100 bp
* 1916 2797: contig of 882 bp in length
* 2798 2897: gap of 100 bp
* 2898 3768: contig of 871 bp in length
* 3769 3868: gap of 100 bp
* 3869 4789: contig of 901 bp in length
* 4790 4869: gap of 100 bp
* 4870 5739: contig of 870 bp in length
* 5740 5839: gap of 100 bp
* 5840 6703: contig of 864 bp in length
* 6704 6803: gap of 100 bp
* 6804 7675: contig of 872 bp in length
* 7676 7775: gap of 100 bp
* 7776 8671: contig of 896 bp in length
* 8672 8771: gap of 100 bp
* 8772 9639: contig of 868 bp in length
* 9640 9739: gap of 100 bp
* 9740 10613: contig of 874 bp in length
* 10614 10713: gap of 100 bp
* 10714 11572: contig of 859 bp in length
* 11573 11672: gap of 100 bp
* 11673 12528: contig of 856 bp in length
* 12529 12628: gap of 100 bp
* 12629 13488: contig of 860 bp in length
* 13489 13588: gap of 100 bp
* 13589 14469: contig of 881 bp in length
* 14470 14569: gap of 100 bp
* 14570 15452: contig of 883 bp in length
* 15453 15552: gap of 100 bp
* 15553 16429: contig of 877 bp in length
* 16430 16529: gap of 100 bp
* 16530 17401: contig of 872 bp in length
* 17402 17501: gap of 100 bp
* 17502 18387: contig of 886 bp in length
* 18388 18487: gap of 100 bp
* 18488 19361: contig of 874 bp in length
* 19362 19461: gap of 100 bp
* 19462 20326: contig of 865 bp in length
* 20327 20436: gap of 100 bp
* 20427 21291: contig of 865 bp in length
* 21292 21391: gap of 100 bp
* 21392 22281: contig of 890 bp in length
* 22282 22381: gap of 100 bp
* 22382 23271: contig of 890 bp in length
* 23272 23371: gap of 100 bp
* 23372 24249: contig of 878 bp in length
* 24250 24349: gap of 100 bp
* 24350 25213: contig of 864 bp in length
* 25214 25313: gap of 100 bp
* 25314 26162: contig of 849 bp in length
* 26163 26262: gap of 100 bp
* 26263 27122: contig of 860 bp in length
* 27123 27222: gap of 100 bp
* 27223 28086: contig of 864 bp in length

```

```
* 28087 28186: gap of 100 bp
* 28187 29073: contig of 887 bp in length
* 29074 29173: gap of 100 bp
* 29174 30055: contig of 882 bp in length
* 30056 30155: gap of 100 bp
* 30156 31136: contig of 981 bp in length
* 31137 31236: gap of 100 bp
* 31237 32116: contig of 880 bp in length
* 32117 32216: gap of 100 bp
* 32217 33100: contig of 884 bp in length
* 33101 33200: gap of 100 bp
* 33201 34070: contig of 870 bp in length
* 34071 34170: gap of 100 bp
* 34171 35169: contig of 999 bp in length
* 35170 35269: gap of 100 bp
* 35270 36164: contig of 895 bp in length
* 36165 36264: gap of 100 bp
* 36265 37135: contig of 871 bp in length
* 37136 37235: gap of 100 bp
* 37236 38163: contig of 928 bp in length
* 38164 38263: gap of 100 bp
* 38264 39164: contig of 901 bp in length
* 39165 39264: gap of 100 bp
* 39265 40131: contig of 867 bp in length
* 40132 40231: gap of 100 bp
* 40232 41124: contig of 893 bp in length
* 41125 41224: gap of 100 bp
* 41225 42122: contig of 898 bp in length
* 42123 42222: gap of 100 bp
* 42223 43129: contig of 907 bp in length
* 43130 43229: gap of 100 bp
* 43230 44109: contig of 880 bp in length
* 44110 44209: gap of 100 bp
* 44210 45090: contig of 881 bp in length
* 45091 45190: gap of 100 bp
* 45191 46090: contig of 900 bp in length
* 46091 46190: gap of 100 bp
* 46191 47072: contig of 882 bp in length
* 47073 47172: gap of 100 bp
* 47173 48053: contig of 881 bp in length
* 48054 48153: gap of 100 bp
* 48154 49033: contig of 880 bp in length
* 49034 49133: gap of 100 bp
* 49134 50009: contig of 876 bp in length
* 50010 50109: gap of 100 bp
* 50110 50981: contig of 872 bp in length
* 50982 51081: gap of 100 bp
* 51082 51957: contig of 876 bp in length
* 51958 52057: gap of 100 bp
* 52058 52945: contig of 888 bp in length
* 52946 53045: gap of 100 bp
* 53046 53932: contig of 887 bp in length.
FEATURES
    source
        1. 53932
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-2LD18"
            /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 14211 a 9672 c 9208 g 15109 t 5732 others
ORIGIN
```

```
Query Match 5.2%; Score 62.6; DB 2: Length 53932;
Best Local Similarity 51.5%; Pred. No. 0.017;
Matches 137; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
```

```
Qy 949 ttacagatgaatttggttactatcatcctctcaaaatggcgcatctattcttta 1008
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29897 TTAATTTAAATAATAATAATAATAATAANANAAATAATAATAAAAAAATTTA 29838
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1009 tttagaagatttgaagcaaatgttatgttgctatgataaatattaataataa 1068
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29837 ATTAATAAATAATAATAATAATAATAATAAAAAAATAATAAAAAAANA 29778
```

```
Qy 1069 aaagatttcgcgatttattagttatttaataagataagaatttcattcttcttta 1128
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29777 AAAAAATTTTAAAAAAATTAATAATNAATNAATNAATNAATNAATNAATNAATA 29718
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1129 aatttatgaacaatttacaggacatttcatttcgagacttggaaaataaagaatgaa 1188
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29717 ATTTTAAATAAATTAATAAAAAAATAATAATTAATAATAATAATAATAATA 29658
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1189 attaatgttaaaaaaataaaaaa 1214
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29657 AAAAAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 29632
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 15
AB019230/c
LOCUS AB019230 74968 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MEB5.
ACCESSION AB019230 BA000014
VERSION AB019230.1 GI:3869069
KEYWORDS Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
SOURCE clone:MEB5.
```

```
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
```

```
REFERENCE
AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
P1 and TAC clones
JOURNAL DNA Res. 7 (2), 131-135 (2000)
MEDLINE 20277480
```

```
REFERENCE
AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
```

```
COMMENT
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MEB5
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi).
```

```
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MIG5 and the 3' clone is MBG14.
```

```
FEATURES
    Location/Qualifiers
        1..74968
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /chromosome="3"
            /clone="MEB5"
            /clone_lib="Mitsui P1"
            join(333..536,644..800,887..1542)
            /note="gene_id:MEB5.1"
            /codon_start=1
            /evidence=not_experimental
CDS
```

/product="purple acid phosphatase-like protein"
/protein_id="BAB02702.1"
/db_xref="GI:9294483"
/translation="MNSGRRLSMATASILLICITFTVVVYVNGELORFIYPAKSDG
SVSFIVIGDMGRGNSVLAYOMCKIGEKIDLDVAVSTGDNFYDNCGLFSEHDNFE
QSFNIIYAPSLQKQNSVIGNHDYNGDREAQLSSVLREIDSRWICLRSFVDAELVE
MFVDTTPFKVEYTEADGHSYDWRVPSRNSYVRLALDLRELVSLKSKARWKIVVGH
HAMRSTIGHGDKELNEELLPIKENGVDYMNHGHCHLQHMDSDESPQIOTSGVGS
KAWRGDINVTINPKLLKYDQGFMSARETHSDAEIVFYDVFGEILHKWTSKQLL
HSSV"
complement(join(1802. .2612,2808. .3118,3284. .3427))
/note="gb|AA02142.1
gene_id:MEB5.2
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02701.1"
/db_xref="GI:9294484"
/translation="MDALTSLVRSPIVPSRTSDNGSGSMFLTAGSGPFTRGSGRLQ
LRLQNASRRSLQSLTITAKTRRSFYVRASASNDASSGSPKPIAPLQLQSPAG
QFLSQTILVSHPHLPAVAQEQLEQTDQSDQNKDSASVPGTDIVLYRRLAEIKEN
ERRTLEETLYALVQKMEANVSLVPSVSPSDPSGRVDYTWPTKVEKLERLHSPMY
EMHNLHALILGRMGDLNSVAQISKLRVGOVYAAVMYGYFLKRVDOQFOLEKTMKI
LFGGDESTISVEQAGCTAYQAASHPEVCAFGAGVSAKGFGESEIKFSRLRSYMS
YMAETLQRTATIRSEAVGIIIEKHTAELFGKPEIIVITPEGTVDSSKDEQIKISFGOMK
RLVLEAVTFGSFLMDVESHVADRYHFVLN"
join(4833. .5114,5267. .5530,5646. .5816,6010. .6122,
6214. .6397,6484. .6593,6687. .6843)
/note="gene_id:MEB5.3"
/codon_start=1
/evidence=not_experimental
/product="senescences-related protein; dihydroorotate
dehydrogenase-like protein"
/protein_id="BAB02704.1"
/db_xref="GI:9294485"
/translation="WASMSFALNRSGLSKTTLSADDPSSRRSFLPPTRVGLKISS
AAESEPDLVTVNGLKMPNFVIGSGPGTNTVKKRADEGWAGVIAKTVSLDASKV
INVTYAKRLTGSGNSAKTDVIGWONIELISDRPLETMKEFERULKREYPRILIAS
VMEYTLAMELIDRVEQTDVALEINFSPCHGPERRMGAAGVDCALLDEVCGWI
NAKATYVWAKMTPNITIDEPARVSLKSGCGIAAINTMSVGMIDMKTLRPEPCVE
GYSTPGYSYKAVPITALAKVMNIAMKMSSEFSEDRSLSGIGVETGYDAAEFLILGS
NVVOVCTGMVHGXYGHVTLCAELDKFMKHNFSTIEEFRGHSLOFTYTHTDLVKRRQK
EAVOEKAEKRGKDKDWTGDFYKTESVSN"
join(8057. .8130,8391. .8430,8506. .8609,8713. .8761,
8841. .8947,9024. .9111,9197. .9400,9480. .9533,9618. .9962)
/note="gene_id:MEB5.4"
/codon_start=1
/evidence=not_experimental
/product="glutamine synthase"
/protein_id="BAB02705.1"
/db_xref="GI:9294486"
/translation="WSLLSDLVNLTDTATGKIIAEYIWIIGSGMDIRSKARTLPQPV
TDPKLPKNYDGSSTGOAGDESEVILYPOAIFKDPFRKGNILVMCDAYTPAGDPI
PNKRHNAKIFSHPDVAKKEPYGIEQVETLMQDVNWPIGMPVGYPGPGPYCG
VGNADKAGDVIADHAKCLYAGIGGINGEWPGEOWEOPVGEISGGQVWVAR
YLLERTEISGVISFDPRVPDGMWAGAHCHNSTKTRNMGGLLEIVIKKAIKGLQLK
HKEHIAAYGEGNERRLTGRHETADINTFSWGVANRGAASVRVGRDTEKEGKGFEDRRP
ASNMDPYVVTSMIAETITLG"
join(12422. .12688,12793. .12861,12959. .13054,13143. .13267,
13534. .13729,13834. .14009,14094. .14223,14304. .14464,
14549. .14645)
/note="gene_id:MEB5.5"
/codon_start=1
/evidence=not_experimental
/product="DnaJ homolog"
/protein_id="BAB02706.1"
/db_xref="GI:9294487"
/translation="MAAMARCALIPINPAHSFRHQFPQPNASFYLLPPLIFSRVRR
FKGISGGRVRRITMAAGTDHYSTLVNVRNATLOETIKSSYRKRLARKYLPDMKNPAGED
KFKITSAAYEVLSDERKRAYDFRFGAGLEGDFNGSQDTPGVDPFDLYSAFFGSDG
FFGGMGSGMGDFDMNKRSLDLDIRYDLRLSFEEAVFGVKREIEVYSLETCDGCGGT
FGKSSNSIKQCSGCDGKGRVMNSQRTPPGIMSQVSCSKGEGEKITDKCRKICNG
RLPARKKMDVVPVPGVSDRATWRIQEGNMKDRSGRAGDLFIVLOVDEKRGIRREGLN

LYSNINIDTDAILGATTKVETVEGSMDLRIPDTPGDTVKLPKRGVDPDTRPSIRG
DHCFFVVKISIPKLLSERKLEVEEFSRLSRSSSTGPTGMLSQSN"
join(16646. .17996,18392. .18984)
/db_xref="gene_id:MEB5.6"
/codon_start=1
/evidence=not_experimental
/product="probable receptor-like protein kinase protein"
/protein_id="BAB02707.1"
/db_xref="GI:9294488"
/translation="MRLFTPSMNSLIFFSILLLSLPLSPSIGDLAAKSAALISFRSA
VGRRTLWDVKOTSPCNMTGVLCDGGRVTALRPGETLSGHIPGEGFNGLTOLRTLUL
RNLGTSLPLDLGSCDLRLYLQGNRESGEIPEVLSLNLRLNLAEENFSEGIS
SGFKNLRLTKLILNNKLSGLLDLSDQFNVNNLNLSIPKSLQKDFSDSFGV
TSLCGKPLVVCNBEQVPSOPISVGNIPTEGSEKKRKKLSSGAIAGIVGCVVG
LSLILMLVLFKGNERTRAIDLATIKHHEVEIPEGAAVEPENRSEVINYSPSA
KVAEYNSGMRKKLVPFGNATKFDLEDLLRASAELVYGDEKLLVYDFMPMGSLALLAV
KRLKVTMDADREFKEKIEVVGAMDHENLPLRAYYVSGDEKLLVYDFMPMGSLALLAV
GNKAGRPPLNWEVRSGLALGARGLDYLHSDQPLSSHGNVAKSNLILTNSHDARSD
TGLAQLVASSTTPNRTGYRAPEVDFPRVSKADVISFGVVLLLELLTGKAPNSVM
NEEGDLARVHSHVAREWRNEVDSELMSIETVSVSEEMAEMLQLGIDICTPQHDPK
RPVMVEVRRIQELRQSGADRVG"
complement(join(20408. .20494,20905. .21059,21182. .21278,
21361. .21417,21496. .21547,21653. .21741,21948. .22033,
22339. .22543,22660. .22752,22834. .22974,23048. .23140,
23292. .23535,23680. .23852,23947. .25299,25378. .25560,
25666. .25730,25899. .26001,26041. .26763,26845. .27042))
/note="gene_id:MEB5.7"
/codon_start=1
/evidence=not_experimental
/product="IRE homolog; protein kinase-like protein"
/protein_id="BAB02708.1"
/db_xref="GI:9294489"
/translation="MTKALNLGPTNSYAFVFLVASSKSLTSKTLKIFORVSASERES
YLFREVRVLEIDNWIYGIWCKSGSSPDSSSPRSVGSNSPIRDKKSKSASKDE
PPTIPFGVGVCKOTIKDGLKKDKGSKQLSEVOAHSIGKNSLSPSSVKKPP
PEVKEGPAFVSPIMASSGLNARIKTRSGPLPOERVENRNDPATNLKMGADGDL
GSGSATSGSGNRRKKEAGSSKLGLEENWDRPSONKSDROSLSPTDTPPRLSPTL
PFSGLQNVASSSGTGLNSDITNNILKWLGIILNRYNRYWQLAVLVLYNLFRVL
LYMSLACRSESSSGSLRNSDCTPENSWENPKESSESPRYQALLRMLAPRKR
FDGIFKSFHELNSGPEWFTEDLLILARSCAMTPTGDFWLQCEGIQVDDLDROEPLP
DIVGLVEKNAEHPWEFTEDLLILARSCAMTPTGDFWLQCEGIQVDDLDROEPLP
GAYKQLTRMLFILTCTRLQDFHKEGWEEQVQLRQSRVLSHTEKIPPSGASRY
SVLKPSTKAYSQEOHGLDKWEDAVRSVPPLAPPENVAIKESPAINDRMSSWK
LPSALKTVKEAPASEQNSKVEPNIVSGRQDDAVALILFPKADSHSHSKH
RINISWYGEQPLISSEESIMCRICEEVPYTHVEDHSRVCTLAKYQKGLSVER
LMAVATDKIAETFRHKDSLAASAPDGMKNSVSHDESDVLSPLSDMSRKGSD
MLDFFPADNSIFMDDLGLMLSCRTFRFGKSDQGMTTSSASMTPRSPPTPRDP
IEQILGKGTFFHDQDIPQMSLADIAKAAADIPGDDQSIPELLSLEDLRYIDRR
KDALVETGTFRIEKLIKEVYHMCLEMDDEKVDLLSTVIDEDADDVDRVLSRTS
PVHPRDTSIDDEFEILKPSRGAFGVFLAKETTGDLFAIKVLKADMTKKNVESI
LAERDLINRVNPFVVPFYSTFCRNLVLMVEYLNNGDLYSLRLNLGLDEEDIVRY
IAEVLLALEYLHSEGVVHRDLKPDNLLIAHDGHKILDTFGLSKVGLINSTDIAAGPA
SETSLDEEESRLAASEQLERKKRSAYTDPYLAPEILLGTHGATADWMSVGIIL
FELIIVGPPENAEHPQIFDNLNLRKIPWHPVEMSAEAHIDIRFLTBDPQRLEGA
RGAAYKOHIFFKDINWDTLAROAAAFVASEAIDTSYFRSRYSNWTSDQFFPEL
VGKLLSLVFGVAISLWNINVAQACEGHAEEFSGIPVDYSFNSFKNLQSLASI
NDLLSKWDEPQQIPHHK"
join(30522. .30606,31683. .31882,32143. .32357,32433. .32523,
32605. .32908,33006. .33066,33143. .33245)
/note="gb|AA09073.1
gene_id:MEB5.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02709.1"
/db_xref="GI:9294490"
/translation="MERDFLGLGSKNSPIITVKEETSESSRDSAPNRMNMFNKNVSA
SSQFLSFRPTQEDRHKSGNYHLPHSGSFMPSADVDTSTKAPYSSVQGVMRMPN
SMOHEETNVAISMFGFQSHHYAPGGRSPMNNNNNQPLVGVPIMAPPISILPPGSI
VGTDTIRSSSKPIGSPAQTLTIFYAGVSVCTDDISPERRAKAMILLAGSSGSPQVFSPP

Query Match 5.1%; Score 62.4; DB 8; Length 74968;
Best Local Similarity 58.7%; Pred. No. 0.019;

Matches 108; Conservative 0; Mismatches 76; Indels 0; Gaps 0;	
Qy 199 gatgaaatcctaaaccgagcagattagatagatactcgaaatcaatcttcaattccggtacta 258	
Db 73408 GGTGAATCCGACACACAGATCAGTTAAGGGTCTTGAAGAGCTTTACCGACAAGGAATA 73349	
Qy 259 ttaacccacacctagagagagattcaaaagaatccggatccggcttcaagaatatggtcaaa 318	
Db 73348 GAATCTCTTCTGCGGACCCACATCCAAACAATCACCGCACAGCTACGACGGTACGGGAAGA 73289	
Qy 319 tcggtgacgcgaacagtggttttactgggttcaaaacccggaaatctcgagcaaaacacaaagc 378	
Db 73288 TAGAAGGCCAAAACGTTTTCTATTGGTTCCAAAACCCACAAGCCCGAGACGCCAGAAAC 73229	
Qy 379 ttcg 382	
Db 73228 GACG 73225	

Search completed: August 27, 2002, 21:34:28
Job time: 4065 sec

THIS PAGE BLANK (USPTO)

QY 171 HNDPSAAQSAFGFDP-VVPPVTEGMAFSTVN-----NGVNLETNE----- 211
 Db 527 ILDKTCTFAEQGGQYDTCIKVIDDAEENVENVQLYNGFVHTGSLBGLSVGDKIIA 586
 QY 212 NFDKI---PAINLY-----GGDNGGNCPPPLTVPLTINOSQEKRDVGLS 254
 Db 587 SFDELRRFPKNNHTGTHILNFALKETLGDNDVDQKGLVAPEKLRDFDSSHKKAVSNEELK 646
 QY 255 GGEDVCDNVYPVRMTVFINEPTEVVSGLFNKAAFGN---DAVLINSFGOPI 304
 Db 647 KVEDICNEIKENLQVYFEIPLDLAKSIDGKRAVEGETYDPVRVVSXGKPI 699

RESULT 2
 US-08-927-219-2
 ; Sequence 2, Application US/08927219
 ; Patent No. 6187533
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Graeme I.
 ; APPLICANT: Yamagata, Kazuya
 ; APPLICANT: Oda, Naohisha
 ; APPLICANT: Kaisaki, Pamela J.
 ; APPLICANT: Furuta, Hiroto
 ; APPLICANT: Horikawa, Yukio
 ; APPLICANT: Menzel, Stephen
 ; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
 ; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/927,219
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/029,679
 ; FILING DATE: 30-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/028,056
 ; FILING DATE: 02-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/025,719
 ; FILING DATE: 10-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wilson, Mark B.
 ; REGISTRATION NUMBER: 37,259
 ; REFERENCE/DOCKET NUMBER: ARCD:272
 ; TELEPHONE: 512/418-3000
 ; TELEFAX: 512/474-7577
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 630 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-927-219-2

Query Match 5.3%; Score 92.5; DB 4; Length 630;
 Best Local Similarity 21.4%; Pred. No. 0.44;

Matches 73; Conservative 43; Mismatches 130; Indels 95; Gaps 19;
 QY 24 QHEIDPTSYMHXNCLNLSFSDRIPDKPR-----WNPKEQIRILSFNS--GTIN 76
 Db 170 OREV-AAQTHAGOGGLIEPTGDELPTKKRRNRKKGWSPASQI-----LFOAYERQKN 223
 QY 77 PREE-----TORIRIRLOEYV-----QTGDANVFYWFQNRKSR--KHKLRV 117
 Db 224 PSKEERETLVEECNRAECIQRGVSPQAQGLGNLVTETVRVYWNFANRKEEAFRHKLAM 283
 QY 118 H-HKSPKMSKKDKTIPSTDADHCFGVNQETGLYVQNNELVVTEPAGFLFP-----VH 171
 Db 284 DYSGPPPGPGPALPA-----HSSPGLPP-----PA--LSFSKVHGVR 321
 QY 172 NDPSAAQSAFGFDPVVVTEGMAFSTVNNGVNLNENFDKIPAINLYGGD---NG 228
 Db 322 GQPATSETA-----EVPSSGGGLVTVSTPLHQVSPGTGLE--PSHLLSTEAKLVSA 371
 QY 229 GGNCFPPLTVPLTINOSQEKRDVGLS-----GGEDVCDNVYPVRMTVFINEP 276
 Db 372 AGGPLPPVST-LTALHSLEQTSPLNQOQPNLIMASLPGYMTIGPGEPASLGPTFTNTGA 430
 QY 277 IEVVSGLFNKAAFGNDVNLNSFGOPILTDFEGVTYOPLO 317
 Db 431 STLVIIGLASTQA---OSVPVINSNGSSL-----TTLQPVQ 462

RESULT 3
 US-08-927-219-4
 ; Sequence 4, Application US/08927219
 ; Patent No. 6187533
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Graeme I.
 ; APPLICANT: Yamagata, Kazuya
 ; APPLICANT: Oda, Naohisha
 ; APPLICANT: Kaisaki, Pamela J.
 ; APPLICANT: Furuta, Hiroto
 ; APPLICANT: Horikawa, Yukio
 ; APPLICANT: Menzel, Stephen
 ; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
 ; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/927,219
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/029,679
 ; FILING DATE: 30-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/028,056
 ; FILING DATE: 02-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/025,719
 ; FILING DATE: 10-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wilson, Mark B.
 ; REGISTRATION NUMBER: 37,259
 ; REFERENCE/DOCKET NUMBER: ARCD:272
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-927-219-4

Query Match 5.3%; Score 92.5; DB 4; Length 630;
Best Local Similarity 21.4%; Pred. No. 0.44;
Matches 73; Conservative 43; Mismatches 130; Indels 95; Gaps 19;

Qy 24 QHEIDTSMYHNSCNLSFSSSSRIDPKPR-----WNPKPEQIRILESIENS--GTIN 76
Db 170 QREV-AQOFTHAGOGGLIEPTGDELPTKKRRNRKMGPSAQI-----LFQAYERQKN 223
Qy 77 PPREE-----IQIRIRLQEG-----QIGDANVFYFQNRKSR--KHKLRV 117
Db 224 PSKEERETLVEECNRAECIQRGVSQAQGLGSLNLTVEVRYNWFANRKEEAFRHKLAM 283
Qy 118 H-HKSPKMSKKDKTVIPSTDADHCFGEFVNQETGLYPVQNNELVTEPAGFLFP-----VH 171
Db 284 DTSGPPPPGPGGPALPA-----HSSPGLPP-----PA--LSPSKVHGVR 321
Qy 172 NDPSAQAQAFGDFGVVPTVEGMAFSTVNGVNLNENFDKIPAINLYGGDG---NG 228
Db 322 QPATSETA-----EVPSSGGPLVTSTPLHQVSPPTGLE--PSHSLLSLEAKLYSA 371
Qy 229 GGNCFPLTVPITINQOEKRDVGLS-----GGEDVDGDNVYPVRMTVFINEMP 276
Db 372 AGGLPPVST-LTALHSLEQTSPLNQOQNOLIMASLPGVMTIGPGEPSALGPTFTNTGA 430
Qy 277 IEVVSGLFNKAAFGNDVNLNFSGQPIILTDGFGVTVYQPLQ 317
Db 431 STLVIGLASTQA---QSVPVINSMGSSL-----TTLQPVQ 462

RESULT 4
US-08-927-219-127
; Sequence 127, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisha
; APPLICANT: Katsaki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/029,679
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,056
FILING DATE: 02-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,719
FILING DATE: 10-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:272
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-927-219-127

Query Match 5.2%; Score 91; DB 4; Length 631;
Best Local Similarity 21.2%; Pred. No. 0.63;
Matches 72; Conservative 43; Mismatches 133; Indels 92; Gaps 18;

Qy 24 QHEIDTSMYHNSCNLSFSSSSRIDPKPR-----WNPKPEQIRILESIENS--GTIN 76
Db 170 QREV-AQOFTHAGOGGLIEPTGDELPTKKRRNRKMGPSAQI-----LFQAYERQKN 223
Qy 77 PPREE-----IQIRIRLQEG-----QIGDANVFYFQNRKSR--KHKLRV 117
Db 224 PSKEERETLVEECNRAECIQRGVSQAQGLGSLNLTVEVRYNWFANRKEEAFRHKLAM 283
Qy 118 H-HKSPKMSKKDKTVIPSTDADHCFGEFVNQETGLYPVQNNELVTEPAGFLFPVH---N 172
Db 284 DTSGPPPPGPGGPALPA-----HSSPGLPP-----PALSPSKVHGVRYG 323
Qy 173 DPSAQAQAFGDFGVVPTVEGMAFSTVNGVNLNENFDKIPAINLYGGDG---NGG 229
Db 324 QPATSETA-----EVPSSGGPLVTSTPLHQVSPPTGLE--PSHSLLSLEAKLYSAA 373
Qy 230 GNCFFPLTVPITINQOEKRDVGLS-----GGEDVDGDNVYPVRMTVFINEMPI 277
Db 374 GGPLPPVST-LTALHSLEQTSPLNQOQNOLIMASLPGVMTIGPGEPSALGPTFTNTGAS 432
Qy 278 EYVSGLFNKAAFGNDVNLNFSGQPIILTDGFGVTVYQPLQ 317
Db 433 TLVIIGLASTQA---QSVPVINSMGSSL-----TTLQPVQ 463

RESULT 5
US-08-591-685-7
; Sequence 7, Application US/08591685
; Patent No. 6083733
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Thermostable xylanases
; NUMBER OF SEQUENCES: 13
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,685
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-591-685-7

Query Match 5.2%; Score 90.5; DB 3; Length 351;
Best Local Similarity 17.6%; Pred. No. 0.29;
Matches 68; Conservative 55; Mismatches 136; Indels 127; Gaps 16;

QY 2 SSSNKNWPSFKSKPCNNHHQHEIDTFSYMHYSNCLSSSFSSDRIPDPKRWPKPE 61
DB 5 S1SDENW-----GQP-----VPDYKLPSCCKYKFKIGVAVPY-RALTNPV 46
QY 62 QIRILESIENSGTINPREIRIRLQRYG---QIGDANFYWFQNRKSRKAKHLRVH 118
DB 47 DVEMIKRHENSIT---PENEMKPESLQYEGGFSFSADEYIDFCKKNISLGRHTLVH 103
QY 119 HKSPKM-----SKDKTVIPSTDADHCFGFVNOETG---LIPVQNNELVTE 162
DB 104 QQTSPFFFTNPETGERLTNSEKDKXILLDRKKHIQTVVGRYKGVYAMDVNEAIDENQ 163
QY 163 PAG-----FLFVHNDPSSAAQSAFGGDFVVPVVTBEGMAFSTVNN- 203
DB 164 PDGYRRSDWNIXGPEYIEKAFIWAHEADPKA---KLFYNDISTEYPYKRGNYLTIKNL 220
QY 204 -----GYNL-----ETNENFDKIPAINLYGGDNGGNGCFPLTVPL 240
DB 221 KAKGVVPVHGVLQCHISLDWPDVSEIETVKLFSRIPGLEIH-----FTEIDISI 270
QY 241 TINQSEKRDVGLSGGEDVDNVPVPMVTFINEMPIEVVSGLFNKAAGNDVAVLNSF 300
DB 271 AKNMTD-----DDAYNRYLLI-----QQAQKLAIFAIDVLKRYNRVTVSTFW 312
QY 301 G-----QPILTDREGVTYQP 315
DB 313 GLKDDYSWLRGDMPLLFDK---DYQP 335

RESULT 6

US-08-591-685-7
; Sequence 4, Application US/08853310
; Patent No. 5948640
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo
; TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppress
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853,310
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 1228.003
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 amino acids

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-853-310-4

Query Match 5.1%; Score 89.5; DB 2; Length 1514;
Best Local Similarity 20.7%; Pred. No. 3.5;
Matches 76; Conservative 41; Mismatches 139; Indels 111; Gaps 14;

QY 26 EIDTPSYMHYSNCL-----SSSFSSDRIPDPKRWNPKEQIRILESIENSGT 74
DB 1077 EVTNPLVMHLLHGNLPLEKVLPPGHRSSRLSSQLP-----LREOSQDRGT 1122
QY 75 -----INPREIRIRIRLQRYGQIGDANFYWFQNRKSRKAKHLRVHHS 121
DB 1123 LOGTGNNRLAARINPGSAQTLESILAQSYG-----ASAGLVRAMASKA 1167
QY 122 PKMSKDKTVIPSTDADHCFGFVNOETGLYPVQNN-----ELVVTEPAGFLFP 169
DB 1168 PAMSOAKIAKMTSLDSQH-----PETELTPSSGNLEIDSKEHLSSLFCEQKEGHSLS 1221
QY 170 VHNDPSSAAQSAFGGDFVVPVVTBEGMAFSTVNGVNLNENFDKIPAINLYGGDGN-- 227
DB 1222 QGSDPGAAP-----GQCLGDHTTSKVPCTSTNVLSFGSEQTDGTLSDQNNAGGHEKKL 1276
QY 228 -GGGNCFFPPLTVPLTINQSOEKRDV-----GLSGG-EDVGDVNVYVPMTVF 271
DB 1277 FPGNTVTTTLCQPSREQTPLPAEVPPVFPSPRKTEPSKNSVSGGVQTTRENRMPKPPPV 1336
QY 272 INEMPTB-----VYSGLFNKAAGNDVAVLNSF-GQPIILD-----EFGVTY-Q 314
DB 1337 ADSIKTEQTELRDPIKADAENRKAAGYSLELVHLOGMPFVVDLPFWKLPREPFGKGFQ 1396
QY 315 PLONGAI 321
DB 1397 PLEPSSI 1403

RESULT 7

US-07-803-636A-2
; Sequence 2, Application US/07803636A
; Patent No. 5422428
; GENERAL INFORMATION:
; APPLICANT: MCGUIRE, TRAVIS C., TERRY F. MCELWAIN, LANCE E. PERRYMAN,
; APPLICANT: WILLIAM C. DAVIS
; TITLE OF INVENTION: IMMUNIZATION AGAINST BABESIOSIS USING
; TITLE OF INVENTION: PURIFIED SURFACE ANTIGENS OF BABESIA BIGEMINA AND SIMILAR
; TITLE OF INVENTION: IMMUNOGENS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 NW 41ST STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FLORIDA
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/803,636A
; FILING DATE: 19911206
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: WA4-059.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-803-636A-2

Query Match 5.1%; Score 88.5; DB 1; Length 480;
Best Local Similarity 21.0%; Pred. No. 0.76;
Matches 42; Conservative 28; Mismatches 73; Indels 57; Gaps

QY 53 KPRWPKPEQIRILEISFNSGTINPPREIQRIRLQEQIGDANFYFQNKRSRAK 112
DB 113 KPLSLPNYQLDAAEMLPREDSDNPAKNEVR-----FMRSSSHGD 155
QY 113 HKLRVHHSKPKSKDKTVPSTADHCFGVNQ---ETGLY-----PVQNE 157
DB 156 Y----HHEFVSLKKN--VVRDPESNDVENFASQFYMTLYKYTLTVDETAAKFFNKL 209
QY 158 LVVTEPAGF-----LFPV-----HNDPSAAQSAFGGDFV---VPVVTEGMAF 198
DB 210 APTTLRFGFGIQALKRLVRSNLPYDLGTHPEATIREIASGYGEYMTQVPAMTSAERF 269
QY 199 STVNGVNLNTNENFDKTPA 218
DB 270 SKMATKTLTVSDYVHLPA 289

RESULT 8
US-08-961-083-4
Sequence 4, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 571 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-4

Query Match 5.0%; Score 88; DB 4; Length 571;
Best Local Similarity 20.1%; Pred. No. 1.1;
Matches 33; Conservative 36; Mismatches 65; Indels 30; Gaps 4;
QY 22 HHQHEIDTPSYMHYNSCNLSsf-----SSDRIPDPKPRWPKPEQIRILEISFNSGTIN 76
DB 42 NOKSSVATPTKQOKVDYNTVPNFVDHPSTVQAIQEQTPVSSTKPTVEQVVVEKPFSTELIN 101
QY 77 PPREEIQRIRLQEQIGDANFYFQNKRSRAKHLRVHHSKPKSKDKTVPSTP 136
DB 102 PRKEEKQ-----SDSQEQLAEHKNLETKKEKISKPKEXTGVNTLN 142
QY 137 A--DHCFGFVNQETGLYPVQNNELVVTEPAGFLFPVHNDPSAAO 178
DB 143 PODEVLSGQLNKPELLYREETWETKI-----DFQEEIQENPDLA 182

RESULT 9
US-08-861-464-4
Sequence 4, Application US/08861464
Patent No. 5874210
GENERAL INFORMATION:
APPLICANT: Guarente, Leonard P.
APPLICANT: Austriaco Jr., Nicanor
APPLICANT: Kennedy, Brian
TITLE OF INVENTION: Genes Determining Cellular Senescence
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/861,464
FILING DATE: 22-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/396,001
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09351
FILING DATE: 15-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,408
FILING DATE: 16-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6408A22
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-861-464-4

Query Match 5.0%; Score 87; DB 2; Length 834;
Best Local Similarity 20.7%; Pred. No. 2.5;

THIS PAGE BLANK (USPTO)

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	100.5	5.7	573	1	YEY8_YEAST	P00285 saccharomyces
2	100	5.7	272	1	HAY5_ARATH	Q02283 arabidopsis
3	99.5	5.7	689	1	YE30_HELPY	P56185 helicobacter
4	99	5.7	559	1	PHO2_YEAST	P07269 saccharomyces
5	98.5	5.6	958	1	SIYAC_YEAST	P04825 saccharomyces
6	98.5	5.6	1159	1	N124_SCHPO	Q09904 schizosaccharomyces
7	97.5	5.6	326	1	PIX2_XENLA	Q09wr3 xenopus laevis
8	97	5.5	251	1	HAY7_ARATH	Q00466 arabidopsis
9	96.5	5.5	305	1	PIX1_XENLA	Q0w751 xenopus laevis
10	95	5.4	692	1	YE30_HELPJ	Q9zj16 helicobacter
11	94.5	5.4	592	1	HN1B_XENLA	Q91739 xenopus laevis
12	94.5	5.4	642	1	MB11_ARATH	Q9sav1 arabidopsis
13	94	5.4	751	1	MYBA_MOUSE	P51960 mus musculus
14	93.5	5.3	280	1	DLX3_AMBME	Q90229 ambystoma mexicanum
15	93.5	5.3	395	1	LHX3_XENLA	P36200 xenopus laevis
16	93.5	5.3	1300	1	PO12_MOUSE	P11369 mus musculus
17	93	5.3	309	1	HXB1_CHICK	P32559 gallus gallus
18	92.5	5.3	192	1	HUNB_DROTA	O46260 drosophila
19	92.5	5.3	407	1	LM06_HUMAN	O43900 homo sapien
20	92.5	5.3	862	1	CUT1_RAT	P53565 rattus norvegicus
21	92	5.3	247	1	DLX6_BRARE	Q98877 brachydanio rerio
22	91.5	5.2	196	1	HUNB_DROAA	O46234 drosophila
23	91	5.2	359	1	KLFB_HUMAN	O95600 homo sapien
24	91	5.2	631	1	HNFA_HUMAN	P20823 homo sapien
25	91	5.2	1531	1	NFT5_HUMAN	O94916 homo sapien
26	90.5	5.2	190	1	HUNB_DROCR	O46236 drosophila
27	90.5	5.2	250	1	HMB1_XENLA	P53773 xenopus laevis
28	90	5.1	195	1	HUNB_DRODA	O46262 drosophila
29	90	5.1	460	1	YMY2_CAEEL	P34479 caenorhabditis elegans
30	89	5.1	605	1	HN1A_XENLA	Q05041 xenopus laevis
31	88.5	5.1	326	1	CRT1_HUMAN	Q15699 homo sapien
32	88.5	5.1	326	1	CRT1_RAT	Q63087 rattus norvegicus
33	88.5	5.1	592	1	EPHD_MYCTU	Q10402 mycobacterium tuberculosis

```

Query Match          5.7%; Score 100.5; DB 1; Length 573;
Best Local Similarity 21.2%; Pred. No. 0.99;
Matches 84; Conservative 46; Mismatches 130; Indels 137; Gaps 20;

Qy      4  SNKNPWSMFKSKPCN-----NNH-----HHQHE-----IDTPSYMH----- 34
      | : | | : : | | | | : | | | | |
Db      92  STRNRPGRYDMMNNMMNNMMNNMMNNMMNNHNTTSDHNAHPQYRCRPNRGRRLSLMTIP 151

Qy      35  --YNCNLSSFSFSDRIDPKPRNKPQEQ--RILESIEN--SGTINPPREEQRI--- 85
      | : | | : : | | : : | | : | | : | | : | | : | | : | | : | |
Db     152  EKYSGSYSLRSSPPTYSNPRVKRELFPFQLRQKMSAQTFPNGENFTPRNQIARLPPS 211

Qy      86  -----RRLQEQYGOIGDANFYVFWQPNKRKRAKH 113

Db     212  STFPDPSSSSLPLTGTGGPSSADNDSIATGNNRNSPOQTAAADA-----NOKSSSES 264
      | : | | : : | | : : | | : : | | : : | | : : | | : : | | : :

```

QY 114 KLRVHHSKPKSKKDKTVPSTADHCFGVNQTGLYVQNNELVTEPAGFLPPVHND 173
 DB 265 PKAIRSNKSKISRFKRKWSKSSADSVESKTKQKRKNRPVVPETISL----- 318
 QY 174 PSAQSAFGFGFVVVTEGMAFSTVNGVNLNENFDKIPAINLYG-----GD 225
 DB 319 -----DQPEIITKQ-----SFSTVNNH-----ET-----AVPSIKDSGIVQELTALGD 356
 QY 226 GNGGNCFPPLTVPLTINOS--QKRDVGL-----SGGEDVG-----DNVYPVRM----- 268
 DB 357 NNR-----IPVLPPSPNRPPLSKRTTKLYCSODSSNEDIAPEEKSTVFLKRLQDEWS 412
 QY 269 TVFINEMPI--EWSGLFNVAAGFNDAVLNSFGQP 303
 DB 413 TVYLKPLTASVPSLSLTITDA-ANSSFINSSISSP 448
 RESULT 2
 HATS_ARATH STANDARD; PRT; 272 AA.
 AC Q02283;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Homeobox-leucine zipper protein HAT5 (HD-zip protein 5) (HD-ZIP protein ATHB-1).
 DE HAT5 OR ATHB-1 OR AT3G01470 OR F4P13.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=91266907; PubMed=1675603;
 RA Ruberti I., Sessa G., Lucchetti S., Morelli G.;
 RT "A novel class of plant proteins containing a homeodomain with a
 RT closely linked leucine zipper motif.";
 RL EMBO J. 10:1787-1791(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
 RA Fartmann B., Vally G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Boutty M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Brandt S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Bangerter S.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
 RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez C., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laude M., Berger-Llauro C., Purnelle B., Masuy D.,
 RA De Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argilou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Waits A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pai G., Miltisger J., Sellers P., Gill J.E., Feldblum T.V.,
 RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Ideasa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana.";

RL Nature 408:820-822(2000).
 RN [3]
 RP SEQUENCE OF 64-163 FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=92237275; PubMed=1349174;
 RA Schena M., Davis R.W.;
 RT "HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3894-3898(1992).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE HD-ZIP FAMILY OF HOMEBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; X58821; CAA41625.1; -;
 DR EMBL; AC009325; AAF01532.1; -;
 DR EMBL; M90416; AAA32816.1; -;
 DR PIR; S16325; S16325.
 DR HSSP; P01366; IAKH.
 DR TRNSPAC; T01474; -;
 DR InterPro; IPR000047; HTH_repressr.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR Homeobox; DNA-binding; Nuclear protein.
 KW DOMAIN 39 63 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 211 243 ASP-RICH (ACIDIC).
 FT DNA_BIND 65 124 HOMEBOX.
 FT DOMAIN 125 160 LEUCINE-ZIPPER.
 SQ SEQUENCE 272 AA; 30933 MW; B85ACCA20EDFA094 CRC64;
 Query Match 5.7%; Score 100; DB 1; Length 272;
 Best Local Similarity 23.1%; Pred. No. 0.41;
 Matches 52; Conservative 35; Mismatches 82; Indels 56; Gaps 11;
 QY 44 FSSDRTPDKPRWNPKEQIRILESIENS-GTINPREETQIRIRLOEQVGIGDANVFY 102
 DB 60 FYDDQDPEKKRLT--TEQVHLLKSFETENKLEPERKTLQAKKLGLQ-----PRQAV 111
 QY 103 WFONRKSRRAKHLRVHHSKPKSKDKTVPST-----DADHCFGVNQTG 149
 DB 112 WFONRRAWK-----TKQLERDYDLLKSTYDQLLSNYDSIVMDNKLSEVSLT- 161
 QY 150 LYPVQNNELVTEPAGFL-FPVHNDPSAQAQSAFGFGFVVVTEGMAFSTVNGV--- 205
 DB 162 -EKLOGKQETANEPPGVPEPNQDPVYINAA-----AIKTEDRLSSGSGVSAVLDD 212
 QY 206 ----NLETNENFDKIPAI-----NLYGGDGGGNGCFPLTVPLT 241
 DB 213 DAPQLDSCDSY--FFSIVPIQDNNASDHNDNRSCFADVFVPTT 255
 RESULT 3
 YE30_HELPY STANDARD; PRT; 689 AA.
 ID YE30_HELPY
 AC P56185;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein HP1430.
 GN HP1430.
 OS Helicobacter pylori (Campylobacter pylori).
 GN Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

```
CC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=9739467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RL Nature 388:539-547(1997).
CC -1- SIMILARITY: BELONGS TO THE UPF0036 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000643; AAD08469.1; -
DR TIGR; HP1430; -
DR InterPro; IPR001279; Beta_lactam_met.
DR InterPro; IPR001587; UPF0036.
DR Pfam; PF00753; Lactamase_B; 1.
DR Pfam; PF02147; UPF0036; 1.
DR PROSITE; PS01292; UPF0036; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 689 AA; 77386 MW; 777919829435B80 CRC64;
Query Match 5.7%; Score 99.5; DB 1; Length 689;
Best Local Similarity 19.9%; Pred. No. 1.5; Mismatches 128; Indels 119; Gaps 18;
Matches 74; Conservative 51;
QY 1 MSSNNKWPMSFKSKPCNNHHOHEIDTPSYM-----NENHENSSENSKADMRAGAFERTNKKRFRNAQKNAEYSN 50
Db 1 MTDNQN-----NENHENSSENSKADMRAGAFERTNKKRFRNAQKNAEYSN 50
QY 38 CNLSSSSSDRIDPKPRNPKPEQIRILESTFNSGTTNPPREIQIRI-----RLQ 90
Db 51 HEASSHHKKEHRPNKPNHHKQKHA-----TRNYAQEELDSNKKVEGVTEILHVN 101
QY 91 EYGOIGDANVFYFONRKSRAKHLRVHKKSP--KWSKKDKTVIPSTADHCFGVNQET 148
Db 102 ERGTG-----FHKEKKGVEANKLQVEHLNPHYKMNLSKASVITPLG----- 147
QY 149 GLYPQNNELVYTEP-----AGLFFVHNDPSAAQSAFGDFGVVPVVTVEGMAFSTV 201
Db 148 GLGEIGGNMVIETPKSAIVIDAGNSFP-----KEGLFGV-DILIP-----DFSVL 192
QY 202 NGVNLETNEFDKTPAHLNLYGGDNGGNN-----CFPPLTVPLINQSQERDVL 253
Db 193 -----HOIKDKTAGIITHAHDHIGATPYLFKEQLFPLYGTPLSL-----GL 235
QY 254 SGG--EDYGDNVYVPMVTVEINEMIEV-----VSLFNKVAFAFGNDVNLINS-FGQPIILT 306
Db 236 IGSKFEDEHGLKKYRFYKIVEKRCPISGVEFIIEWHITHSIDSSALAIQKAGTIHT 295
QY 307 DEFGVYQPLQN 318
Db 296 GDFKIDHTPDVN 307
RESULT 4
```

```
PHO2_YEAST
ID PHO2_YEAST STANDARD; PRT; 559 AA.
AC P07269;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
GN PHO2 OR BAS2 OR GRF10 OR YDL106C OR D2350.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87146334; PubMed=3029672;
RA Sengstag C., Hinnen A.;
RT "The sequence of the Saccharomyces cerevisiae gene PHO2 codes for a
RL regulatory protein with unusual aminoacid composition.";
RN [2]
RP SEQUENCE FROM N.A.
RA Berben G.;
RL Submitted (JAN-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97051597; PubMed=8896274;
RA Salz J.E., Buitrago M.J., Garcia R., Revuelta J.L., del Rey F.;
RT "The sequence of a 20.3 kb DNA fragment from the left arm of
RL PHO2, POL3 and DUN1 genes, and six new open reading frames.";
RN [4]
RP STRUCTURE, EXPRESSION, AND FUNCTION.
RX MEDLINE=89006272; PubMed=3049251;
RA Berben G., Legrain M., Hilger F.;
RT "Studies on the structure, expression and function of the yeast
RL regulatory gene PHO2.";
RL Gene 66:307-312(1988).
RN [5]
RP HOMEBOX DOMAIN.
RX MEDLINE=88210458; PubMed=2896548;
RA Buerklin T.R.;
RL Cell 53:339-340(1988).
CC -1- FUNCTION: REGULATOR IN PHOSPHATE METABOLISM AND ACTS AS A
CC DEREPRESSOR OF ANOTHER CENTRAL REGULATOR PHO5. BINDS TO THE
CC TRP4, HIS4, AND CYC1 PROMOTERS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: TWO REGIONS OF STRONG HOMOLOGY TO PHO4 ARE FOUND.
CC ALSO SIMILAR TO N-CRASSA NUC1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05062; CAA28729.1; -
DR EMBL; M22259; AAA34866.1; -
DR EMBL; X95644; CAA64906.1; -
DR EMBL; Z74154; CAA98673.1; -
DR PIR; A25872; A25872.
DR HSSP; P02836; 1HDD.
DR TRANSFAC; T00689; -
DR SGD; S0002264; GRF10.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
```

DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT DOMAIN 23 GLN-RICH (INVOLVED IN TRANSCRIPTIONAL
FT ACTIVATION) (POTENTIAL).
FT DNA_BIND 77 136 HOMEBOX.
SQ SEQUENCE 559 AA; 63390 MW; BE20E396D6AA0281 CRC64;

Query Match 5.7%; Score 99; DB 1; Length 559;
Best Local Similarity 26.7%; Pred. No. 1.3;
Matches 40; Conservative 24; Mismatches 59; Indels 28; Gaps 7;

QY 20 NHHQHEIDTPSYMYHNSNLSSTSSDRIPDPKPRWPKPQIRILSIFNSGRIINPPR 79
Db 53 NLEHDD-----QHTNDMSASSNDSGPPQRPRKTRAKGEALDVLKRKFE---INPTP 102
QY 80 ERIQIRIRLOEYGOIGDANFYFQNRKSRRAKHLRVHHSKPKMSKKDKTVPSTDA-D 138
Db 103 SLVERKKI--SOLIGMPEKNRINWQNRRAKKRKK---QHGS-----NKDTIPSSQSRD 151
QY 139 HCFGFVNQETGLYPVQNNELVYTPAGFLF 168
Db 152 IANDYDRGST-----DNNLVTTTSTSSIF 175

RESULT 5
SYAC_YEAST
ID SYAC_YEAST STANDARD; PRT; 958 AA.
AC P40825;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Alanyl-tRNA synthetase, cytoplasmic (EC 6.1.1.7) (Alanine--tRNA
ligase) (AlaRS).
GN ALA1 OR YOR335C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP Rippmaster T.L., Schimmel P.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051586; PubMed=8896263;
RA Parle-Mcdermott A.G., Hand N.J., Goulding S.G., Wolfe K.H.;
RT "Sequence of 29 kb around the PDR10 locus on the right arm of
Saccharomyces cerevisiae chromosome XV: similarity to part of
chromosome I.";
RL Yeast 12:999-1004(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U18672; AAC49007.1; -
CC ENBL; Z49821; CAA89980.1; -
CC EMBL; Z75243; CAA99658.1; -
CC SGD; S0005862; ALA1.
CC InterPro; IPR002106; AA_trna_ligase_II.
CC InterPro; IPR003156; DHHA1.
CC InterPro; IPR002318; tRNA-synt_2c.
CC Pfam; PF02272; DHHA1; 1.
CC Pfam; PF01411; tRNA-synt_2c; 1.

DR PRINTS; PR00980; TRNASYNTHALA.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT CONFLICT 136 R -> S (IN REF. 1).
FT CONFLICT 136 KDO -> RTK (IN REF. 1).
FT CONFLICT 465 467 FE -> LQ (IN REF. 1).
FT CONFLICT 840 841
SQ SEQUENCE 958 AA; 107277 MW; 50FD31C2E2D40F32 CRC64;

Query Match 5.6%; Score 98.5; DB 1; Length 958;
Best Local Similarity 19.7%; Pred. No. 2.8;
Matches 49; Conservative 43; Mismatches 120; Indels 37; Gaps 8;

QY 79 REEIQRIRLOEYGOIGDANFYFQNRKSRRAKHK---LRVHHSKPKMSKKDKTVPST 135
Db 465 KQDQLKLNLVHSELNDKVV-----PKTNDFFKYSANVEGTILKLDGTNFEVDIT 518
QY 136 DADHCFGFVNQETGLYPVQNNELVWTEPA-----GFLFPVHNDPSAAQ 178
Db 519 EPGKKYGIILDKTCFYAEOGGQYDTGKIVIDDAEENNVQLYNGFYF---HTGSLEE 575
QY 179 SAFGFGDFVVPVTEEGMAFSTVNNQVLETNENFDKIPAINLYGGDGGNGCFFPLTV 238
Db 576 GKLSVGDKII-ASFDELRRFPKNN---HTGTHILNFALKETGLNDVDQKGLVAPEKL 630
QY 239 PLTIQSQEKRDVGLSGGEDVDGNYVPVMTVFINEMPIEVVSGLFNVKAAFGN---DAV 295
Db 631 RFDSSHKKAVSNEELKKVEDICNEIKENLVQFYKEIPLDLAKSIDGVRVFGETTPDPV 690
QY 296 LINSFGQPI 304
Db 691 RVSVGKPI 699

RESULT 6
N124_SCHPO
ID N124_SCHPO STANDARD; PRT; 1159 AA.
AC Q09904;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nucleoporin nup124 (Nuclear pore protein nup124).
GN NUP124 OR SPAC30D11.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99340273; PubMed=10409764;
RA Balasundaram D., Benedik M.J., Morpew M., Dang V.-D., Levin H.L.;
RT "Nup124p is a nuclear pore factor of Schizosaccharomycetes pombe that
is important for nuclear import and activity of retrotransposon Tfl.";
RL Mol. Cell. Biol. 19:5768-5784(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Pearson D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT. IN S. POMBE IT IS
REQUIRED FOR THE NUCLEAR LOCALIZATION OF RETROTRANSPON TFL.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -!- DOMAIN: CONTAINS 11 X-F-X-F-G REPEATS.
CC -!- SIMILARITY: SOME, TO YEAST NUCLEOPORINS NUP1 AND NUP2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF280406; AAF90179.1; -
DR EMBL; Z67961; CA91890.1; -
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 270 274 POLY-ARG.
SQ SEQUENCE 1159 AA; 123974 MW; 852F84BA2E0619A3 CRC64;

Query Match 5.6%; Score 98.5; DB 1; Length 1159;
Best Local Similarity 20.9%; Pred. No. 3.6;
Matches 63; Conservative 40; Mismatches 120; Indels 79; Gaps 14;

QY 10 SMFKSKPCNNHHQHEDTPTSYHMYNCLNLSFSSDRIPDPKPRWPKPQIRILESI 69
DB 444 SQFSSPKETTRKSEVPES-----PSKEIKSHSFVPEFKPEKTEAT----- 489
QY 70 FNSGTINPPREE-----IQIRIRLQE-----YQIGDANVFYFQNRKRSRAKHK-- 114
DB 490 -TDKKNVPKEEPTATADVQTNRLKENEPKPTFAQLPSKT-----QETPSITENKPS 543
QY 115 --LRVHKSPKMSKDKT-VTPSTDADHCFGVNQETGLYPVQNNELVVTEPAGLFPVH 171
DB 544 FFSQSPKRETEKKNAPASAPASTSGFSFG-----GFAPKTLKEEETKAPTENFSLN 597
QY 172 NDPAAQSA-----FGFG-DFVVPV-----VTEEGMAFSTVNGVNLNENFDKIP 217
DB 598 NASSTQDTTKTLQFNFGSSGKPGKPTNSFNNDKKTSSENGLASSTV-----ASESKPSAP 650
QY 218 AINLYGGDGGGNGCFPLTVPLTINQSQEKRDV-----GLSGGDVGDNVYVPRMT 269
DB 651 ESKPSSGFGNTAGSS-----PFSNLTKESEVPPTNSFSAKKGKDEANDLSAKAST 704
QY 270 VF 271
DB 705 PF 706

RESULT 7
PIX2_XENLA STANDARD; PRT; 326 AA.
AC Q9PWR3; Q93396; Q9YHA4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Plutitary homeobox 2 (xPtx2).
GN PITY2 OR PITY2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM PITY2A).
RX MEDLINE=98370612; PubMed=9707115;
RA Ryan A.K., Blumberg B., Rodriguez-Esteban C., Yonei-Tamura S.,
RA Tamura K., Tsukui T., de la Pena J., Sabbagh W., Greenwald J.,
RA Choe S., Norris D.P., Robertson E.J., Evans R.M., Rosenfeld M.G.,
RA Izpisua-Belmonte J.-C.;
RT "Pitx2 determines left-right asymmetry of internal organs in
RT vertebrates.";
RL Nature 394:545-551(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS PITY2A AND PITY2B).
RX MEDLINE=99146886; PubMed=10021341;
RA Campione M., Steinbeisser H., Schweickert A., Delasler K.,
RA van Biebler F., Lowe L.A., Nowotzsch S., Viebahn C., Haffter P.,
RA Kuehn M.R., Blum M.;
RT "The homeobox gene pitx2: mediator of asymmetric left-right signaling
RT in vertebrate heart and gut looping.";
RL Development 126:1225-1234(1999).

CC -!- FUNCTION: May play an important role in development and
CC maintenance of anterior structures and mediate left-right
CC asymmetry. Could play a role at the interface of lateral plate
CC signaling and heart and gut morphogenesis.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; PITY2A (shown here) and PITY2B;
CC are produced by alternative splicing.
CC -!- DEVELOPMENTAL STAGE: Asymmetrically expressed in the left lateral
CC plate mesoderm, tubular heart and early gut tube.
CC -!- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC "BICOID" SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 OAR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF077767; AAC29426.1; -
DR EMBL; AJ005786; CAA06696.1; -
DR EMBL; AJ005787; CAA06697.1; -
DR HSSP; P06601; 1FJL.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox.1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Alternative splicing.
KW DNA_BIND 94 153
FT DOMAIN 288 301
FT DOMAIN 294 298
FT DOMAIN 24 33
FT VARSPLIC 1 70

CONFLICT 78 78
FT CONFLICT 83 83
FT CONFLICT 158 158
FT CONFLICT 190 190
FT CONFLICT 306 307
FT CONFLICT 310 310
FT CONFLICT 323 323
SQ SEQUENCE 326 AA; 36339 MW; 74366F0E4FB48E73 CRC64;

Query Match 5.6%; Score 97.5; DB 1; Length 326;
Best Local Similarity 22.2%; Pred. No. 0.83;
Matches 58; Conservative 39; Mismatches 83; Indels 81; Gaps 11;

QY 18 NNNHHQHEDTPTSYHMYN-----NC-----NLSFSSDR-----I 49
DB 26 HHHHHHHQHTVTLVSMASLQSGECKSRLEVHTISDTSSPDADKDSHQTKNEDSST 85
QY 50 PDP-----KPRWPKPEQIRILESI--NSGTINPPREEIQIRIRLQYEQIGDANV 100
DB 86 DDPKSKKRQRORTHTFTSQQLQLEATFORNYPDMSTREEI-----AVTNLTARV 138
QY 101 FYWFONRKSRAKHLRVHHSKMSKKDKTVPSTPDADHCFGVNQETGL-----YVP 153
DB 139 RYWFKNR--RAKWRKRNRQQAELCKN-----GFGPQFNGLMQPYDDMYPS 182
QY 154 QN-----NELVVTPEAGFLPPVHNDPSAAQSAFGFGDFVVPVVTVEEGNAFTV-- 201
DB 183 YSYNNWAAGKLTASLSTKSPFFFNMMNVNPLUSSQSSPNSISSMSWSSGMVPSAVTG 242

RESULT	10
YE3O_HELPJ	
ID	YE3O_HELPJ STANDARD; PRT; 692 AA.
AC	Q9ZJ16;
DT	16-OCT-2001 (Rel. 40, Created)
DF	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypothetical protein JHP1323.
GN	JHP1323.
OS	Helicobacter pylori J99 (Campylobacter pylori J99).
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC	Helicobacter.
NCBI_TaxID	=85963;
[1]	
RN	SEQUENCE FROM N.A.
RX	MEDLINE=99120557; PubMed=9923682;
RA	Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA	Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA	Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA	Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA	Trust T.J.;
RT	"Genomic sequence comparison of two unrelated isolates of the human
RT	gastric pathogen Helicobacter pylori.";

```

RL Nature 397:176-180(1999).
CC -|- SIMILARITY: BELONGS TO THE UPF0036 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001555; AAD06901.1; -.
DR InterPro: IPR001279; Beta_lactam_met.
DR InterPro: IPR001587; UPF0036.
DR Pfam: PF00753; Lactamase_B.1.
DR Pfam: PF02147; UPF0036; 1.
DR PROSITE: PS01292; UPF0036; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 692 AA; 77676 MW; 2E318AB33AFA45F6 CRC64;
-----
Query Match 5.4%; Score 95; DB 1; Length 692;
Best Local Similarity 19.6%; Pred. No. 3.6;
Matches 73; Conservative 54; Mismatches 128; Indels 118; Gaps 19;
-----
Qy 1 MSSNKNWPMFKSKPCNNHH-----OHEIDTPSYMHY-----S 36
| : : : | | : : | : : :
Db 1 MTDNNQ-----NENHSENSKDHHEARAGAFERFTNKKRFRENAQKNAESS 50
-----
Qy 37 NCNLSSFSFSDRIPDPKRWPKPEQIRILSINSGTINPREEIQIRI-----RL 89
| : : : | | : : | : : : | : : :
Db 51 NHETLSHHKKHRRPNKKPNNNHKPKH-----SQKTRNAKELDNNKVEGVTEILHV 103
-----
Qy 90 QEYGOIGDANFYWFQNRKSAKRLRVHHKSP--KMSKKDKTVIPSTDADHCFGVNQE 147
| : : : | : : : | : : : | : : :
Db 104 NERTGLG---FHRELKKGVEANKIQVHNLPHYKMLNSKASVKITPLG----- 150
-----
Qy 148 TGLVPVQNELLVTEP-----AGFLPVPVNDPSAASAGFGDFVVPVTEEGMAFST 200
| : : : | : : : | : : : | : : :
Db 151 -GLGEIGGNMVIETPKSAIVIDAGMSPP-----KEGLFGV-DILIP-----DFS 194
-----
Qy 201 VNNGVNLETNEFDKIPAINL-YGGDGGNGGN-----CFPLTPVLTINOSQEKRDVG 252
| : : : | : : : | : : : | : : :
Db 195 L-----HQIKQKIAGIITHAHEDHIGATPYLFKELQFLYGTPLSL-----G 237
-----
Qy 253 LSGG--EDVGDNVYPMRTVTFINEMPIEV---VSGLEFNKAAFGNDAVLINS-FGQPIL 305
| : : : | : : : | : : : | : : :
Db 238 LIGSKFDEHGLKLYRSYFKIVEKRCPISGEFIEMWIIHTSIIDSSALAIOTKAGTIH 297
-----
Qy 306 TDEFGVTVQPLQN 318
| : : : | : : : | : : :
Db 298 TGDFEKIDHTPVDN 310
-----
RESULT 11
HN1B_XENLA STANDARD; PRT; 592 AA.
ID HN1B_XENLA
AC Q91739; O05042; P79924;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hepatocyte nuclear factor 1-alpha B (HNF-1A-B) (HNF-1-alpha B) (LFB1)
DE (XLFBLB).
GN LFB1B.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_taxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94085786; PubMed=8262383;

```


Search completed: August 27, 2002, 20:29:46
Job time: 433 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2002, 20:17:47 ; Search time 19.07 Seconds
(without alignments)
1637.600 Million cell updates/sec

Title: US-09-787-737-2
Perfect score: 1751
Sequence: 1 MSSSNKNWSPMFKSPKNN.....TDEFGVTYQLONGAIYILI 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	467	26.7	321	2 G84750	probable homeodoma
2	217.5	12.4	295	2 E84425	probable homeodoma
3	200	11.4	292	2 T00829	wuschel protein -
4	188	10.7	293	2 D84558	probable homeodoma
5	175	10.0	197	2 G86339	protein F2010.20 [
6	168.5	9.6	244	2 A84687	probable homeodoma
7	167	9.5	268	2 T04660	hypothetical prote
8	166.5	9.5	249	2 F96511	hypothetical prote
9	106	6.1	3097	2 T28635	glutamate synthase
10	104.5	6.0	956	2 T40953	hypothetical prote
11	104	5.9	751	1 I49497	transforming prote
12	100.5	5.7	573	2 S50661	hypothetical prote
13	100	5.7	272	2 S16325	homeotic protein A
14	99.5	5.7	689	2 F64698	conserved hypothet
15	99	5.7	559	2 A25872	transcription regu
16	98.5	5.6	958	2 S2065	alanine--tRNA liga
17	98.5	5.6	1159	2 S62562	probable nuclear p
18	98	5.6	468	2 T45476	heat-shock protein
19	97	5.5	251	2 T52370	homeobox protein H
20	97	5.5	251	2 T49950	homeobox-leucine z
21	96.5	5.5	594	2 J70766	transcription fact
22	96	5.5	873	2 J64863	homeobox protein z
23	95	5.4	420	2 C96995	uncharacterized co
24	95	5.4	692	2 F71821	hypothetical prote
25	95	5.4	771	2 T38616	hypothetical prote
26	95	5.4	938	2 T39006	related to yeast z
27	95	5.4	1072	2 G5851	probable hemolysin
28	95	5.4	13055	2 T16580	hypothetical prote
29	94.5	5.4	642	2 A96560	hypothetical prote

30	93.5	5.3	362	2 T24046	hypothetical prote
31	93.5	5.3	395	1 S38821	homeotic protein 1
32	93.5	5.3	848	2 T00372	hypothetical prote
33	93.5	5.3	1281	1 GNMSLL	retrovirus-related
34	93	5.3	309	1 A60096	homeotic protein h
35	93	5.3	647	2 JQ2149	B west mating prot
36	93	5.3	1184	2 T41515	coiled coil protei
37	92.5	5.3	570	4 B44282	retrovirus-related
38	92.5	5.3	862	2 B53689	homeotic protein C
39	92.5	5.3	971	2 T03181	hypothetical prote
40	92	5.3	508	2 E71620	hypothetical prote
41	92	5.3	1262	2 T25168	hypothetical prote
42	91.5	5.2	962	2 T22459	hypothetical prote
43	91	5.2	542	2 S39608	transcription fact
44	91	5.2	605	2 A48115	transcription fact
45	91	5.2	631	1 A36749	transcription fact

ALIGNMENTS

RESULT 1
G84750
Probable homeodomain transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84750
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: G84750
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <STO>
A:Cross-references: GB:AE002093; NID:g2253582; PIDN:AAC69146.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g33880
A:Map position: 2

Query Match 26.7%; Score 467; DB 2; Length 321;
Best Local Similarity 45.5%; Pred. No. 6.6e-31;
Matches 111; Conservative 28; Mismatches 51; Indels 54; Gaps 11;

QY	1	MSSSNKNWSPMFKSPKNNHHQHHEIDTPSYNHYSNCNLSSFS--DRIPDPKPRW	56
Db	1	MASSNRHWPSPKSKP--HPHQWHDINSP--LLPSASHRSSPSSGCEVERSPKPRW	56
QY	57	NPKPEQIRILESFNSGTINPREEIQIRIRLOEQIGDANFYWFQNRKSRKHKLR	116
Db	57	NPKPEQIRILEAFNSGMVNPREEIRIRRAQLQEQYGVGDANFYWFQNRKSRKHKLR	116
QY	117	VHH-----KSPKMSKKDKTVIP-----STDADHCFVFNQETG	149
Db	117	LLHNHSHKSLPQTQPQPQPSASSSSSSSSSKTKPKRKNKNTNLSLG-GSQMMG	175
QY	150	LYPVQNNELVTEPAGFLFPVHN-----DPSAAQSAFGF--GDPV---VPVVTTEGMAFST	200
Db	176	MFP-----PEPA-FLFPVSTVGFGTIVSSQLGFLSGDMIEQOKPAPTCTGLLSE	226
QY	201	VNNG 204	
Db	227	IMNG 230	

RESULT 2
E84425
Probable homeodomain transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: G86339

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-197 <STO>

A:Cross-references: GB:AE005172; NID:g8886931; PIDN:AAF80617.1; GSPDB:GN00141

C:Genetics:

A:Gene: F2D10.20

A:Map position: 1

Query Match 10.0%; Score 175; DB 2; Length 197;
Best Local Similarity 34.9%; Pred. No. 4.1e-07;
Matches 38; Conservative 22; Mismatches 35; Indels 14; Gaps 3;

Qy 53 KPRWNPKEQIRILESIENSGTINPPREIQRIRIRLOEYGOIGDANFYWFQNRKSRK 112

Db 77 RPRWPTTTQLILENIYKESGTPNPRRIKEITMELSEHGOIMEKNYHWFQNRKSRK 136

Qy 113 HKLRVHHKSP-----KMSKKDKTIVIPSTD-----ADHCFGFVNQETGLYP 152

Db 137 RK-----QPTTTITSSQADDAATTEERGCGDSDSGFESYEHILFP 180

RESULT 6

A84687

probable homeodomain transcription factor [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: A84687

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: A84687

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-244 <STO>

A:Cross-references: GB:AE002093; NID:g4580396; PIDN:AAD24374.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g28610

A:Map position: 2

Query Match 9.68%; Score 168.5; DB 2; Length 244;
Best Local Similarity 44.9%; Pred. No. 1.9e-06;
Matches 35; Conservative 11; Mismatches 27; Indels 5; Gaps 1;

Qy 50 PDKPRWNPKEQIRILESIENSGTINPPREIQRIRIRLOEYGOIGDANFYWFQNRKS 109

Db 3 PVASTRWCTPEQMLEENRSGIRTPNAVQIQOITAHAFYKRIEGKNVYWFQNHKA 62

Qy 110 RAKHKLVRHHKSPKSKK 127

Db 63 RDRQKLR-----KKLAKQ 75

RESULT 7

T04660

hypothetical protein F8D20.60 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-May-1999

C:Accession: T04660

R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Jesse, T.; Heijnen, L.; Vos, P.; Mewes,

submitted to the Protein Sequence Database, July 1998

A:Reference number: Z15381

A:Accession: T04660

A:Molecule type: DNA

A:Residues: 1-268 <BEV>

A:Cross-references: EMBL:AL031135

A:Experimental source: cultivar Columbia; BAC clone F8D20

C:Genetics:

A:Map position: 4

A:Introns: 75/1; 241/1

A:Note: F8D20.60

Query Match 9.5%; Score 167; DB 2; Length 268;
Best Local Similarity 46.8%; Pred. No. 2.9e-06;
Matches 29; Conservative 15; Mismatches 18; Indels 0; Gaps 0;

Qy 53 KPRWNPKEQIRILESIENSGTINPPREIQRIRIRLOEYGOIGDANFYWFQNRKSRK 112

Db 97 RQRWPTTVQLILRIFDQGTGTPSKQIKDITELSHQGOIABQNVYWNFNRRSRK 156

Qy 113 HK 114

Db 157 RK 158

RESULT 8

F96511

hypothetical protein F2G19.11 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F96511

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

ansen, N.F.; Hughes, B.; Huizar, L.

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F96511

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-249 <STO>

A:Cross-references: GB:AE005173; NID:g11321765; PIDN:AAG34242.1; GSPDB:GN00141

C:Genetics:

A:Gene: F2G19.11

A:Map position: 1

Query Match 9.5%; Score 166.5; DB 2; Length 249;
Best Local Similarity 25.4%; Pred. No. 2.9e-06;
Matches 50; Conservative 26; Mismatches 48; Indels 73; Gaps 7;

Qy 24 QHEIDTPSYMHYSCNLSFSSSDRIPDPKPRWNPKEQIRILESIENSGTINPPREEIQ 83

Db 73 EHKRPPPHOTHPGG-----TRWNTPQOIGILEMLYKGGMRTPAQQIE 116

Qy 84 RIRIRLOEYGOIGDANFYWFQNRKSRKHKLRVHHK-----SPKMSKKD 128

Db 117 HITLQKYGKIEGKNVYWFQNHKARERQKQRNNLSLSQCSFTTGTGVNPSVTMKT 176

Qy 129 KTVIPSTDAD-----HCFGFVNQETGLYPVN--NELVVTEPA 164

Db 177 RT---SSSLIDIMREPVEKEELVEENYKRTCRSGFENLE-----IENRRKNKSNSTMAT 228

Qy 165 GF-----LFPVH 171

Db 229 TFNKIIDNWTLELFPPLH 245

RESULT 9

T28635

glutamate synthase (NADH) (EC 1.4.1.14) - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

Db 102 ERGTLG-----FHKELKGVKAEANKIQVEHLNPHYKMNLSKASVKITPLG----- 147

Qy 149 GLYPVQNNELVTEP-----AGFLFPVHNDPSAAQSAFGDFVVPVVTVEEGMAFSTV 201

Db 148 GLGEIGGNMVIETPKSAIVIDAGHSFP-----KEGLFGV-DILIP-----DFSVL 192

Qy 202 NNGVNLFTNENFDKIPAINL-YGGDNGGGN-----CFPPLTVPLTTNQSQKRDVGL 253

Db 193 -----HQIKDKIAGIITIHAHEDHIGATPYLFKELQFPFYGTPLSL-----GL 235

Qy 254 SGG--EDVCDVNPVVRMTVFINEMPLEV-----VSGLFNVKAAFGNDVAVLINS-FGQPIILT 306

Db 236 IGSKDFEHLKKYRSYFKIVEKRCPLISVGFEFTIEWIHTHSIDLSALAIQNAGTIIHT 295

Qy 307 DEFVGYTQPLQN 318

Db 296 GDFKIDHTPVDN 307

RESULT 15

A25872

N:transcription regulator GRF10 - yeast (Saccharomyces cerevisiae)

N:Alternate names: acid phosphatase synthesis regulatory protein; protein D2350; protein

C:Species: Saccharomyces cerevisiae

C>Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000

C:Accession: A25872; S67648; S19636; S67412; S72100

R:Sengstag, C.; Hinnen, A.

Nucleic Acids Res. 15, 233-246, 1987

A:Title: The sequence of the Saccharomyces cerevisiae gene PHO2 codes for a regulatory p

A:Reference number: A25872; MUID:87146334

A:Accession: A25872

A:Molecule type: DNA

A:Residues: 1-559 <SEN>

A:Cross-references: GB:M24613; NID:g4143; PIDN:CAA28729.1; PID:g4144

R:Ballesta, J.P.G.; Remacha, M.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.M.; B

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67629

A:Accession: S67648

A:Molecule type: DNA

A:Residues: 1-559 <BAL>

A:Cross-references: EMBL:274154; NID:g1431149; PIDN:CAA98673.1; PID:e253225; PID:g143115

A:Experimental source: strain S288C

R:McCarthy, B.J.; Creasy, C.L.; Bergman, L.W.

Nucleic Acids Res. 19, 3463, 1991

A:Title: Molecular analysis of a temperature sensitive allele of the PHO2 gene of Saccha

A:Reference number: S19636; MUID:91288241

A:Accession: S19636

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 65-156, 'E', 158-310, 'N', 312-559 <MCC>

A:Cross-references: EMBL:X54293; NID:g4146; PIDN:CAA38192.1; PID:g4147

A:Experimental source: strain GG100-14D

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1990

R:Bozkovic, J.; Saliz, J.E.; Soler-Mira, A.; Garcia-Cantalejo, J.; Revuelta, J.L.; Jimine

submitted to the EMBL Data Library, February 1996

A:Reference number: S67406

A:Accession: S67412

A:Molecule type: DNA

A:Residues: 1-559 <BOS>

A:Cross-references: EMBL:X95644; NID:g1199535; PIDN:CAA64906.1; PID:e223187; PID:g119954

R:Saliz, J.E.; Buitrago, M.J.; Garcia, R.; Revuelta, J.L.; del Rey, F.

Yeast 12, 1077-1084, 1996

A:Title: The sequence of a 20.3 kb DNA fragment from the left arm of Saccharomyces cerev

A:Reference number: S72094; MUID:97051597

A:Accession: S72100

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-559 <SAI>

A:Cross-references: EMBL:X95644; NID:g1199535; PIDN:CAA64906.1; PID:e223187; PID:g119954

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996

C:Genetics:

A:Gene: SGD:GRF10; PHO2; BAS2

A:Cross-references: MIPS:YDL106c; SGD:S0002264

A:Map position: 4L

C:Function:

A:Description: required for expression of phosphate pathway and other genes; acts as

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:78-134/Domain: homeobox homology <HOX>

Query Match 5.7%; Score 99; DB 2; Length 559;

Best Local Similarity 26.7%; Pred. No. 3.2;

Matches 40; Conservative 24; Mismatches 58; Indels 28; Gaps 7;

THIS PAGE BLANK (USPTO)